DOCUMENT RESUME

ED 182 071 RC 011 792

TITLE	Basic Skills Learning Centers Evaluation. Final Report 1 October 1976 - 30 September 1979.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	Community Services Administration (DHEW), Washington, D.C.
PUB DATE	1 Oct 76
NOTE	138p.
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Achievement; Basic Reading; *Basic Skills; Bilingual Education; Critical Thinking; Elementary Education; Inservice Teacher Education; Instructional Materials; Language Skills; *Learning Laboratories; *Program Development; *Program Effectiveness; Program Evaluation; Remedial Instruction; Remedial Mathematics; Research Methodology; Research Problems; *Rural Schools; Sequential Programs; Socioeconomic Background; Spanish Speaking; Tables (Data); Technical Assistance
IDENTIFIERS	*Basic Skills Learning Centers Project; United States (West)

ABSTRACT

Detailed program descriptions and discussion of research methodology are included in this independent evaluation of the Basic Skills Learning Centers (BSLC) Projects implemented by Southwest Educational Development Laboratory (SEDL) and Southwest Regional Laboratory (SWRL) and designed to improve basic reading and math instruction in nonurban elementary schools. Following an introductory statement of the scope and purpose of the evaluation, Chapter II describes the SEDL and SWRL-BSLC programs including materials, staff training and technical assistance components, and target populations (monolingual English-speaking students for SEDL, Spanish-speaking students for SWRL). Chapter III covers initial planning and development of the evaluation format, rationales for design decisions, and explanations of the methodological approach. Chapter IV contains extensive tabular and narrative analyses of test and survey data gathered during the evaluation. Program implementation and student test performance are discussed separately for SEDL and SWRL programs. Significant academic gains are shown for students in both programs. Chapter V discusses implications of the BSLC Project for future educational efforts, emphasizing the special features of bilingual programs. The appendices (RC 011 793) contain a wide array of supporting materials. (JH)

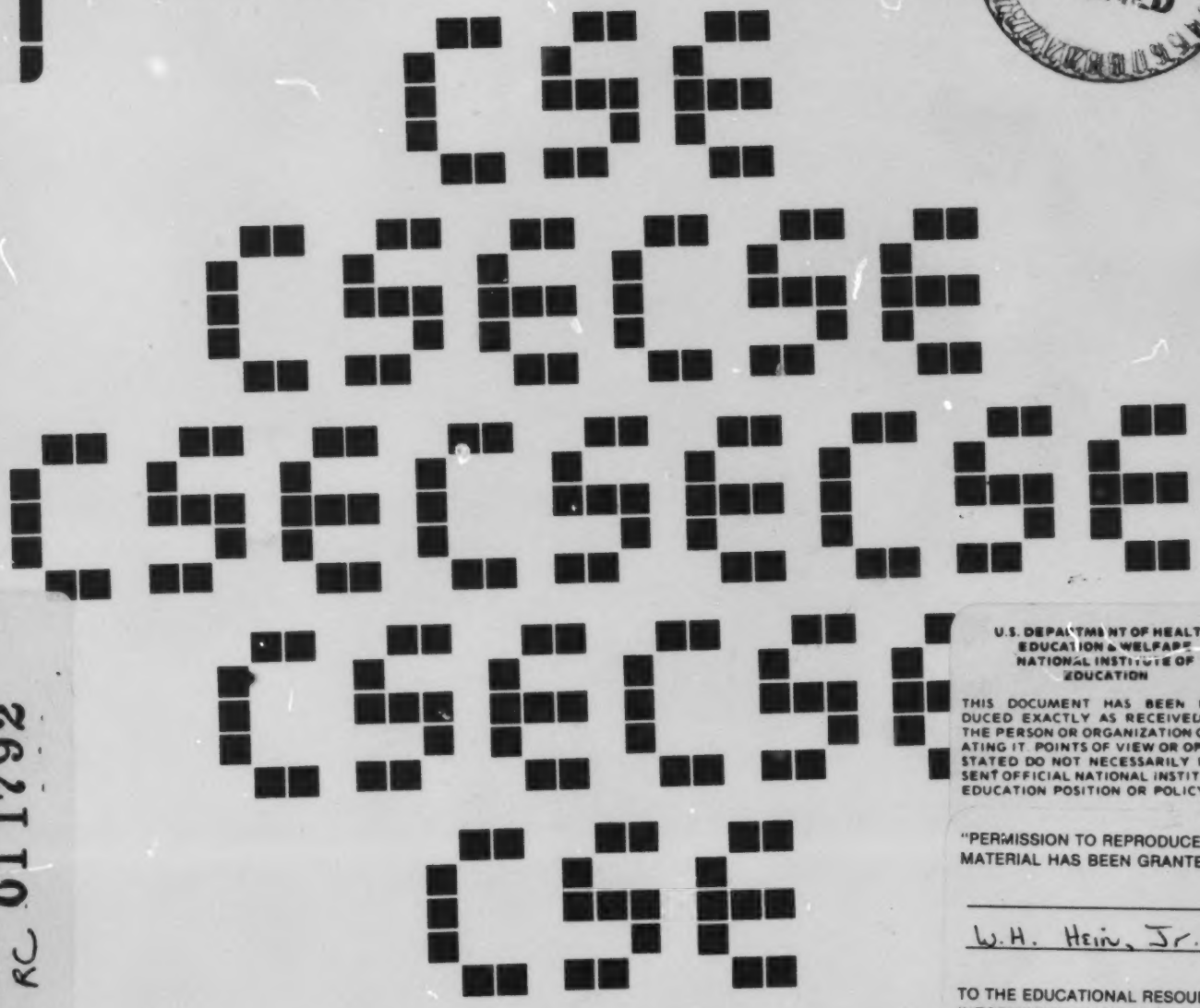
****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Center for the Study of Evaluation

UCLA Graduate School of Education
Los Angeles, California 90024

RC

ED182071



RC 011792

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W.H. Hein, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Center for the Study of Evaluation

The Center for the Study of Evaluation (CSE) is a university-based research and development center established in 1966 by the Department of Health, Education, and Welfare. The mission of the Center has been to conduct research, development, training, and dissemination activities for the purpose of improving the planning and evaluation of education and other social action programs. More specifically, CSE activities include: guidance for educators and researchers in the areas of tests and measures and in educational decision making; development of easy to use evaluation procedures and materials in the form of manuals, workshops and kits; and evaluation of projects for federal, state, local, and university agencies requesting technical assistance in evaluation. The work of CSE during its first decade has clarified the ways in which systematic evaluation can be helpful in the improvement of many kinds of programs.

Information about CSE and its publications may be obtained by writing to:

Director, Public Information
Center for the Study of Evaluation
UCLA Graduate School of Education
Los Angeles, California 90024

FINAL REPORT

1 October 1976 - 30 September 1979

BASIC SKILLS LEARNING CENTERS EVALUATION

Submitted to

COMMUNITY SERVICES ADMINISTRATION

Center for the Study of Evaluation
University of California
Los Angeles, California

TABLE OF CONTENTS

Executive Summary i

Chapter I: Introduction 1

Chapter II: Descriptions of the SEDL and SWRL Programs 5
- SEDL Program Description 6
- SWRL Program Description 10

Chapter III: Evaluation, Design and Method 13
- Selection of the Evaluation Design 13
- Instrumentation 24
  - SEDL-BSLC Implementation Survey 24
  - Measurement of SEDL Program Implementation 25
  - Measurement of SWRL Program Implementation 36
- Data Collection and Sampling 46
  - SEDL-BSLC Implementation Survey 46
  - SEDL-BSLC Testing Instruments 48
  - SWRL-BSLC Testing Instruments 54
  - Field Visits 57

Chapter IV: Results 58
- SEDL Results 58
  - SEDL Implementation Survey Results 58
  - SEDL Student Test Data 72
- SWRL Results 89
  - SWRL Implementation Survey 89
  - SWRL Student Test Data 109

Chapter V: Conclusions and Recommendations 119
- Conclusions and Recommendations Concerning the SEDL-BSLC Program 119
- Conclusions and Recommendations Concerning the SWRL-BSLC Program 121
- The BSLC Evaluation: Its Implications for the Field 123

CHAPER I: INTRODUCTION

This document is intended to satisfy the reporting requirements for the Center for the Study of Evaluation (CSE) evaluation of the Basic Skills Learning Centers Program (BSLC), under contract with the Community Services Administration. Presented herein are the results of three years of evaluation activity aimed at documenting the nature and effects of the BSLC Program.

Background for the BSLC Initiative

In response to legislative mandate to provide educational services to disadvantaged students in need of remediation in the basic skill areas, CSA, in collaboration with the National Institute of Education, evolved a plan whereby specific developement efforts underway at the NIE's regional laboratories would be adapted for use as "Basic Skills Learning Centers". After review of resources available at the labs, the Southwest Regional Laboratory (SWRL) and the Southwest Educational Development Laboratory (SEDL) were chosen to implement the service component of the program. CSE was selected as an appropriate agency to conduct an independent evaluation of the impact of the BSLC program. Following these selections, sole source procurements were approved for SWRL and SEDL to prepare and implement the curriculum components, and for CSE to plan and implement the called for evaluation.

The SEDL and SWRL-BSLC Programs

Within the targeted service area there was sufficient latitude for the two labs to develop quite distinct operationalizations of the BSLC concept.

The SEDL-BSLC program is bilingual and is intended for use with children in grades kindergarten through third, who may be limited English proficient. The program is designed to provide instruction in Spanish, English, or in both languages. SEDL program components consist of bilingual oral language and

reading (BOLaR) and bilingual continuous progress mathematic (BCPM). The materials are culturally based for Mexican-American students.

The SWRL-BSLC program is designed for use with monolingual English speaking children, and does not make explicit provision for ethnolinguistic differences among students. The program is intended to supplementary remedial instruction for reading and mathematics skills normally attained in grades K - 6. The SWRL program components consist of Basic Reading Skills (BRS), Decoding Skills I and II (DS I-II) and Reading Development Skills I-III (RDS I-III), for reading, and Basic Mathematics Skills I-VI (BMS I-VI) for mathematics. The SWRL materials focus on direct practice on sequenced skills.

The CSE Evaluation

CSE's initial choice of orientation for the CSA inquiry focused on what is called "implementation" evaluation. In such an evaluation, the question of focus is the extent to which a program gets used as intended by the developers or seriously modified in the course of its dissemination. Such studies are usually precursive to studies of program effects, in that some evidence that the service is being delivered or in research terms, that the "treatment", is in place should be accrued before inferences about its effects can be justified. CSE's opinion was that in mounting two large, multi-district, complex instructional systems, as the SWRL and SEDL programs undoubtedly were, this verification of the delivery of services was extremely important. Over a period of considerable discussion, it became clear that CSA expected that student performance data where essential to the evaluation effort.

In an attempt to be responsive to this perspective, CSE developed a control group design in which students in the program and out of the program would be contrasted in terms of their achievement. When logistical difficulties

suggested that this design should be modified, and approximation was developed. After a series of iterations involving the participating educational laboratories and CSA project monitors, the final strategy, not involving the collection of comparative data, emerged.

During the 1977-78 and 1978-79 school years CSE gathered data that would reflect student performance with regard to program specific objectives and national performance standards for both SWRL and SEDL. In addition, information concerning program features, installation, usage patterns and teacher characteristics was obtained. CSE implemented an evaluation responsive to the issue of whether students participating in the SEDL and SWRL programs had improved their skills in the areas of reading and mathematics as a result of having received basic skills instruction in those areas. The following general questions guided the evaluation effort:

- What is the nature of the population being served?
- How do the Basic Skills Learning Centers function with regard to installation and usage patterns?
- How has the performance of participating students changed as a result of instructional intervention?
- What is the relationship between specific program components (i.e., reading and mathematics) and specific student outcomes?

The evaluation attempts to answer these questions by analysing the implementation and student outcome data that was gathered during the two years of program operation for the SEDL and SWRL Basic Skills projects.

Overview of the Report

The remainder of the report is divided into four chapters. In Chapter II we provide a detailed description of the SEDL and SWRL-BSLC programs, including the nature of the materials, training and technical assistance

components. Chapter III presents the features of the evaluation, including initial planning and development activities, rationales for design decisions, explanations of the methodological approach, instrumentation, data collection and sampling activities. Chapter IV contains the results of our analyses of test and survey data gathered during the evaluation. The results are presented separately for the SEDL and SWRL programs; within each descriptions of program implementation and student test performance are provided, followed by analyses of the effects of the BSLC program on student achievement outcomes. The final chapter, Chapter V, presents a discussion of the findings of the evaluation, including conclusions about program effects and the limitations of those conclusions, ending with a statement about the implications of the BSLC program and evaluation for future educational improvement efforts. The appendix contains a wide array of supporting materials, ranging from psychometric data on the instruments to summaries of field notes taken during visits to participating BSLC sites.

CHAPTER II: DESCRIPTIONS OF THE SEDL AND SWRL PROGRAMS

The Basic Skills Learning Center (BSLC) programs, developed by both the Southwest Regional Laboratory (SWRL) and the Southwest Educational Development Laboratory (SEDL) provide remedial instruction in reading and mathematics for elementary school students. The SEDL program also includes attention to thinking and reasoning skills, although this component was not included in the present study.

Both laboratories provide three primary services to implement their respective BSLC programs:

1. Learning Center Resources - the materials intended for students
2. Operations Resources - the materials intended for school personnel to help them implement the program
3. Technical Assistance - direct service from laboratory personnel

The programs of the two laboratories are similar in that they feature objectives-based materials in kit form, utilize school district coordinators as primary liasons to school site personnel and provide a telephone "hot-line" for technical assistance needs. However, there are striking differences in the approach to remedial instruction, and the student populations served, as described in greater detail below. SEDL's programs are multi-media core curricula that provide total instructional environments for teaching students essential primary grade reading and mathematics skillls; the programs are aimed predominantly at students of Hispanic background and include instruction in English and Spanish. In contrast, SWRL's materials are designed for monolingual English students, and are intended as supplementary to the core instructional program; in the main, the materials provide direct practice in the reading and mathematics skills expected at grades one through six.

This difference in scope, i.e., core vs. supplementary instruction, has important consequences both for the costs of the respective programs, the number of students served, and the amount of operations services required and provided. Because SEDL's multi-media, core program is clearly more costly per learning center than is the SWRL program, the SWRL program could be implemented in 145 districts and 458 schools, while the SEDL program could be implemented in 60 districts and approximately 200 schools. It has been estimated that the SWRL program serves approximately 96,000 students, while the SEDL program serves approximately 15,000. In addition, because SEDL's program intends to provide a total instructional environment, it requires significant teacher inservice training and staff development, and program implementation is more time and labor intensive.

SEDL PROGRAM DESCRIPTION

The SEDL Basic Skills program is Spanish/English bilingual and is intended for use with students primarily of Hispanic background who may have limited proficiency in English, limited proficiency in Spanish or bilingual speakers.

The program is directed toward the prevention and remediation of learning problems specific to reading, mathematics, and cognitive skills (thinking and reasoning skills) for the kindergarten through third grade levels. The program does not attempt to group students who have lower achievement levels into special classes for the purpose of remedial instruction, but rather aims to circumvent the need for later remediation by providing an effective instructional program for all students from the start.

Designed predominantly for students in first and second grade classrooms, the SEDL program may also be used to provide remedial instruction to older students whose level of achievement is commensurate with first or second grade.

Although the program advocates no single approach in instruction, it is intended as an individualized core curriculum which can be adapted to local school and administrative environments.

Learning Center Resources

The SEDL program offers three Learning Center Resources for students: Bilingual Oral Language and Reading (BOLaR), Bilingual Continuous Progressive Mathematics (BCPM), and Thinking and Reasoning (T & R).

The BOLaR program addresses basic communication skills in English and Spanish. Its goals are to develop listening, speaking, reading and writing skills in a curriculum that is culturally based for Mexican-American students. Oral language activities in both languages are planned to accommodate the kinds of learning differences that exist for students who lack proficient speaking skills. The oral language components are intended to provide the basis for subsequent instruction in English and/or Spanish reading. The BOLaR program includes four components: Spanish oral language, English oral language, Spanish reading, and English reading. The Spanish oral language component provides second language instruction to native English speakers, remedial instruction to native Spanish speakers and mainline instruction to students participating in bilingual programs.

The English oral language component presents second language instruction to native Spanish speakers. Students are taught through small group activities that focus on familiar domains. Instruction is phonologically based.

The Spanish reading component proyides basic instruction in Spanish to native English speakers and to students who are participating in bilingual programs. Spanish reading is introduced using a phonetically based instructional strategy. This component focuses on developing pre-reading skills and emphasizes visual-auditory discrimination, vowels and consonants, comprehension, and syllabification.

The English reading component provided basic instruction in English for native Spanish speakers or first reading experience for native English speakers. This component consists of three sequences which incorporate a language-experience approach, a sight-word approach, visual-auditory discrimination tasks and introductory phonics.

The BCPM program instructs students in first and second grade basic mathematics skills. The program is designed in an aural-visual mode and can be implemented in English or Spanish, independent of the students reading ability. The BCPM program provides initial instruction for students with limited proficiency in English and remedial instruction for students who have not yet mastered specific mathematics skills. The program contains 93 objectives arranged into 12 mathematical areas at two progressive levels of difficulty. The system of diagnosis designed for the BCPM component allows students to enter the program at their own level of skills accomplishment.

T & R is a process curriculum designed to teach students analytic thinking and problem solving skills. This component is written in English. Students learn to develop reasoning skills through the use of SEDL developed games and other activities that they may initiate on their own. The T & R component does not depend on a student's reading ability. It purports to teach the kinds of motivational skills necessary for independent problem solving and to present concepts that may transfer into other curricular areas.

SEDL employs an eclectic instructional approach within and across program components. Teachers and aides use a wide range of methods, techniques and activities to teach individual skill/content areas. Though SEDL contracted with CSA to implement their materials for the purpose of remediation, each of the components could serve as the principal source of instruction at the intended grade levels. SEDL estimates that the reading and mathematics components may

take longer than two years of core instruction to complete. Apart from the DCPM program, the SEDL materials do not provide specific assessment measures for the individual program components.

The SEDL Learning Center concept is flexible. All program components are to be used as an integrated package, but the organizational pattern allows for personalized and individualized instruction focusing on specific skill areas, as necessary. The system allows students to enter a program of instruction at his or her own level and to progress at an individualized pace in specific mathematics and reading skills. Students may proceed in sequence through program materials or individual components from one of several starting points, depending on the teacher's assessment of the students initial skill level. Learning centers are not restricted to any specific grouping strategies, physical settings or logistical procedures, though guides for instructional management are provided.

Operations Resources

The operations resources furnished by the SEDL program combine both the instructional environments and implementation features. These resources include information systems and assistance in using learning centers resources. The information systems consist of materials and procedures that specify the skills addressed in each of the program components and guidelines to the teachers and aides for assessing students' skill accomplishments. Handbooks detail requirements, such as a description of the desired physical environment of the classroom, grouping strategies, and the timing and integration of the program components.

SEDL offers training and implementation pre- and inservice workshops for teachers, aides and other school personnel involved with the Basic Skills Program. Training for classroom users of the bilingual reading strand takes

three days, and a Spanish speaking teacher or aide is needed for effective instruction. Training for implementation of the bilingual mathematics and the cognitive strands requires approximately a day and a half each. Training is provided in the use of specific learning center resources and in the integration of the Basic Skills Learning Center Program into the general school environment. A goal of this effort is to bridge the gap between basic skills objectives, kindergarten through third grade school curriculum, bilingual program objectives and the overall organizational structure of the school. A coordinator is designated at each site to act as the liaison person between SEDL and the schools involved.

SWRL PROGRAM DESCRIPTION

The SWRL-BSLC program provides supplementary remedial instruction for reading and mathematics skills normally attained in kindergarten through grade 6. The program is intended for monolingual English speaking students, and does not make special provisions for the needs of ethnic, or linguistic minorities, although the program currently serves a significant number of Spanish-surnamed, native-American, Asian-American and Black students.

In contrast to the SEDL approach, the SWRL materials focus on direct practice on sequenced skills. The materials are not designed to contain general background, motivational, or enrichment activities. Because they furnish a minimum of direct instruction enough to prepare students for the practice exercises, they are more appropriate as supplementary instructional materials adaptable to any curriculum or mode of instruction.

The SWRL materials were developed in repeated cycles of empirical testing and revisions, and the materials themselves reflect this view of systematic instruction. Within the program, opportunities are given for pre- and post-instruction testing and general monitoring of student progress. Results of

these assessments by the site are sent by participating sites to SWRL which prepares school and district level reports from the data.

Learning Center Resources

The SWRL material consist of six kits in reading (Basic Reading Skills, Decoding Skills I and II, and Reading Development Skills I-III) and six kits in mathematics (Basic Mathematics Skills I-VI). Each of these Learning Center resources is objectives-based, and reflects a structured sequence of skills derived from an analysis of the skill domain. Each kit contains practice materials, with removable, reusable exercise cards for student use. Also included for each student are highly organized periodic skill checks, and structured guidelines for assigning individual students to instruction.

The format of the kits, with individual practice exercises, allows teachers considerable flexibility in individually assigning materials during instruction. The materials are amenable to any instructional approach or curriculum, and in contrast to the SEDL program, do not require additional classroom staffing beyond the teacher. The materials, then, are quite flexible and adaptable for remedial purposes. They may be used to reduce pressure on students by allowing them to work at their own pace on materials that fit their needs, and are appropriate for strengthening prerequisite skills that are necessary for on-grade level instruction.

Operations Resources

The operations resources provided by the SWRL-BSLC program are planned to help teachers and district personnel in program implementation. These include a Learning Center Operations kit which accompanies each set of materials, a District Guide, and orientation and pre-program planning. The Learning Center Operations kit contains a guide for using the materials, record sheets for reporting student progress and successes, pupil assignment folders, and placement aides for determining the objectives individual students need to start working

on. This operations kit, in conjunction with staff orientation and training of less than a day is intended to provide teachers with all the procedural resources necessary for implementing the program. District Guides are provided to district coordinators to give them an overview of activities, their tasks and responsibilities for the year, planning aids, and suggestions for responding to school staff questions about the program.

Technical Assistance

As with the SEDL program, a telephone hotline to SWRL is available to program users to answer questions and concerns. Also similar to SEDL, SWRL staff are available for on-site visitation to provide, where needed, additional support for implementation.

SWRL also provides, as technical assistance, the BSLC Information Service, which summarizes at the school and district-levels pupils' accomplishments in the BSLCs.

CHAPTER III: EVALUATION, DESIGN AND METHOD

In this chapter we summarize the basic features of the evaluation design and methodology employed for the CSE evaluation of the Basic Skills Learning Centers program. The SEDL and SWRL-BSLC program were in operation for two school years, and during each of these years CSE was active in monitoring the progress of samples of participating students, and the activities and reactions of the teachers responsible for this progress. The general goal of CSE's evaluation was to produce information that provided a sound and clear picture of how and to what extent CSA's support of educational improvement brought about positive changes in the students reached by the SEDL and SWRL programs. Our efforts to achieve this goal took somewhat different forms for the evaluations of the two BSLC programs, while at the same time retaining a set of common elements to ensure responsiveness to the terms of our contract with CSA.

We begin this chapter with a discussion of the features common to the evaluations of both programs and of the development of the evaluation design. This section followed by a summary of the design features that were unique to the two different programs. After this introduction to the evaluation design, we provide a description of the intrumentation for each aspect of the design and conclude with an account of the implementation of the evaluation (i.e., data collection and sampling), by program and year of operation.

Selection of the Evaluation Design

Both the central thrust of CSA's involvement in education and the terms of CSE's evaluation contract are concerned with achievement in the basic skills, i.e., reading and mathematics. The specific language surrounding

the program speaks of children achieving below the national norms in the basic skill areas. For the evaluation, this concern translates into a requirement to document student outcomes of the BSLC program in a way that allows comparison with national standards. On the other hand, the CSA mandate to provide remediation in the basic skills to low SES, rural children requires that the design be sensitive to the achievement patterns of this special group of students, often underrepresented in national norming studies. While ideally the evaluation would establish whether or not the BSLC programs advance the achievement status of target students relative to students in the population at large, two kinds of difficulties arose that diminished the likelihood and desirability of making these normative comparisons. First the technical requirements for a rigorous comparison of program effects against a "no treatment expectation" would have been difficult to meet within the scope of effort implied by the funding level for the evaluation. Some approximation to the ideal was needed. Second, the substance of the two BSLC programs and the manner in which the programs were implemented were such that experimental-type comparison of the programs to equivalent control groups would be logistically complex and difficult to attach meaningful interpretations to. Given the need, then to select an appropriate design possessing methodological integrity and responsiveness to the specific programs being examined, considerable effort was devoted during the planning year (1976-77) to the development of an appropriate design and instrumentation battery (see the appendix for a summary of the first year activities).

In addition to technical considerations and responsiveness to programmatic effects specific to the two labs' BSLC materials, selection of the

evaluation design was influenced by the labs' activities during the planning year. Proposed designs and instrumentation were reviewed by all participants in the evaluation, SWRL, SEDL, CSE and its advisor, NIE. The eventual design thus evolved from interaction among the agencies, regarding their information needs, the potential methodological problems, and the time and resource parameters of the evaluation.

Returning to the question of technical rigor in evaluating student performance, the issue resolves into a discussion of the feasibility of obtaining some form of data indicative of how students would have done had they not been exposed to the BSLC program. This is often thought to boil down into whether or not the evaluator gathers data on "contol" students not participating in the program. Early in the history of the program the joint decision was reached not to include a control group in the evaluation design, apparently ruling out the possibility for a clear contrast between treatment and control groups. Prior to the implementation of the BSLC programs, a potentially promising alternative to a separate control group, the "control item," was proposed by Chris Walker of CSE. Several lines of reasoning led to the decision to abandon the control item approach (including basic lack of information as to its soundness and its appropriateness to the BSLC evaluation). The participants in the evaluation (CSE, CSE, SEDL and SWRL) came to basic agreement on the role treatment group comparisons would play in the evaluation.

In the absence of a control group for comparison data, it became necessary to seek another procedure for substantiating statements about the programs' impact on students. Given that data would be available only

on participating students, the question shifted to one of planning a data collection strategy that would yield information interpretable in terms of progress vis-a-vis national student achievement standards. The logic behind any such strategy could be of two sorts. First, if students were tested with a norm-referenced achievement test before and after their remediation, and data reflecting the amount of instruction received during the interval between testing occasions were also collected, one could relate amounts of gain shown by students to the amount of instruction they received. By this reasoning, the analysis of the data would divide the gain in test scores into two parts, one that is predictable on the basis of knowledge of the students' pretest scores alone, i.e., the part of the gain that would be expected even if the program were not implemented, and another part that was apparently related to exposure to the program. If this latter part of the gain were substantial, it would be evidence that the program increased students' scores beyond the level we would have predicted knowing only how the students performed before entering the program. Note, however, that the analysis plan does not imply that all students' scores were improved by the program; rather, it suggests that the more program-related instruction students receive, the higher their scores will be. This is essentially the rationale underlying student outcome analyses.

The second possible rationale for interpreting data from a single standard does not require the collection of information about program implementation. This approach is exemplified by the models adopted by the Office of Education for use in local evaluations of Title I projects (Tallmadge and Wood, 1976). It involves making an assumption about how students in the program would have progressed if they were not exposed to the program

(or if the program were no more beneficial than what it was to supplement or replace). In particular, the model assumes that, without intervention, the students would maintain their percentile rank relative to the national norms on the achievement test. For example, if the group of students targeted for the program were scoring at the 25th percentile, i.e., less well than 75% of the nation's students, before entering the program, the model predicts that after exposure to the program they will still be scoring at the 25th percentile, if the program is ineffective; if their percentile rank is advanced after receiving the program treatment, the presumption is that the experience was beneficial for them. A variant of the Office of Education Title I evaluation model, by making an additional assumption, allows the evaluator to administer a norm-referenced achievement test on only one occasion if an appropriate criterion-referenced test is given both before and after exposure to the program.

Both approaches - the program-outcome analysis and the Title I evaluation model - depend on assumptions that may be unrealistic. The first approach assumes that student growth follows a straight line, without program intervention. In the second case, the assumption is made that, without remediation, students remain stable in their achievement level, when in fact such students often tend to fall further behind. It is not clear that the first problem produces any bias in favor of or against the program being analyzed; furthermore, even the most rigorous of evaluation methods in practice are subject to the same criticism. The Title I model, however, may well bias the analysis toward underestimating the program's impact, since a program that maintains the students' relative standing could in fact be an effective intervention, in the face of predicted performance decrement. Both approaches make a third assumption related to

the sensitivity of norm-referenced achievement measures to the effects of specific instructional interventions.

As mentioned above, the strategy used in our analyses of BSLC student outcomes involves estimating the amount of achievement gain that is associated with increments in exposure to BSLC instruction, other things being equal. In the case of outcome scores derived from a nationally normed test, this means that the analysis will provide an indication of the extent to which students' standing in comparison to national achievement levels could be improved by participation in the BSLC program. We use the words "could be" because the analysis does not assume that all students have been exposed to influential amounts of BSLC instruction. By imagining a variable that measured how much BSLC instruction students received, we can show how the proposed analysis strategy relates to the traditional two group treatment/control analysis. In the traditional model, one attempts to determine whether being in the treatment group is associated with higher outcome scores, on the average, than are obtained in the control group. This is equivalent to the extreme case in our instructional exposure analysis, where each student received either no BSLC instruction, i.e., a score of zero on the imaginary variable, or some presumably adequate amount of instruction.

Figure 1 portrays the analogy between the treatment-control analysis and the instructional exposure analysis. In the left frame, the zero for the control group and "X" for the treatment group, and the vertical axis shows the amount of gain exhibited by the two groups. The right frame depicts the single group case where more fine-grained data are available on the amount of the treatment each student in the group

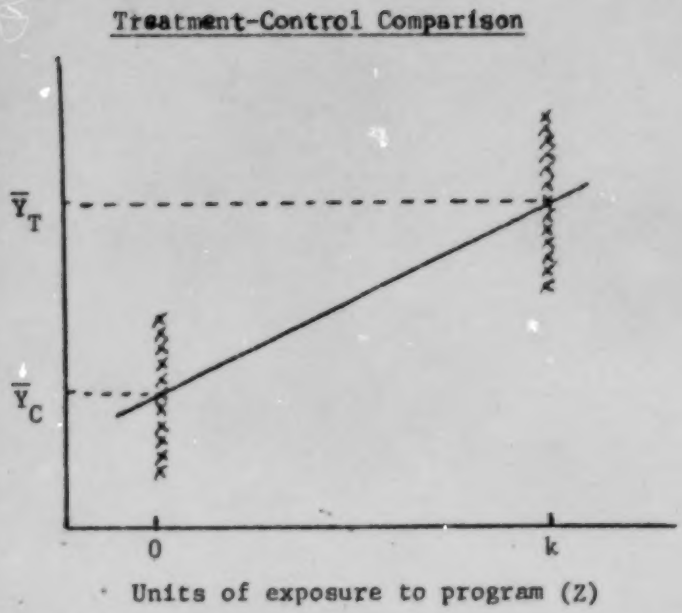
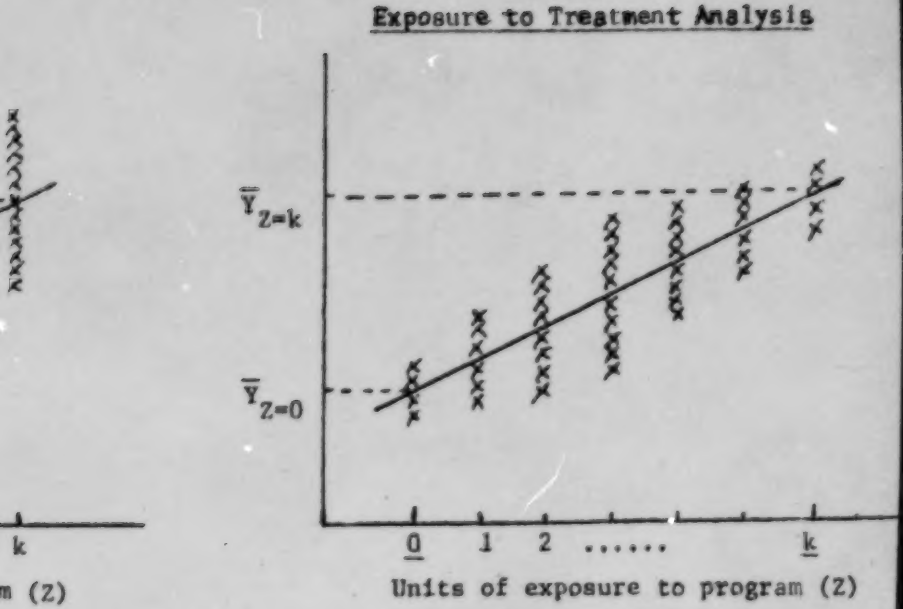


Figure 1. Relationship between treatment-control group exposure to instruction on treatment group o



ent-control group analysis and an analysis of the effects of treatment group outcomes.

received. If the amount of treatment ranged from zero units to "k" units, as the right frame indicates, we can see that the average amount of gain for those students in the group with zero and k units of exposure to treatment is the same as for the two groups in the left frame. This equivalence is further demonstrated by the fact that we can fit the same "regression" line to the data in both frames. The trade-off between the two strategies can be seen to involve collecting data on students not receiving the program versus collecting data on the amounts of the treatment only those students allegedly participating in the program actually were exposed to.

One final point about the strategy for analyzing student gain needs to be made. The planned analysis requires three pieces of information to be available, a measure of the students' performance prior to beginning the program, an index of exposure to the program, and a measure of students' performance after the program. In order to interpret the relationship between outcome and exposure to treatment in the manner proposed above, it is necessary that the "pre-measure" used in the analysis be a completely adequate summary of what the students knew about the material to be taught, before actually being exposed to this material. When that portion of achievement at the end of the program that can be predicted from the pre-measure is removed, the remainder of the posttest gain may legitimately be attributed the amount of the treatment students received. Nothing in the logic of the analysis demands that the pre-measure be precisely the same test used for measuring program outcomes: instead, all that is required is that the pre-measure be an adequate summary of students' mastery of the relevant content prior to the onset of BSLC instruction. Thus, even though the goal of the analysis is to describe the impact of the program on students' norm-referenced achievement, the pretest used in the analysis need not be a norm-referenced test.

In summary, the major common element in the approaches to the evaluation of the SEDL and SWRL programs is the effort to document the impact of the BSLC program on participating students, in terms of norm-referenced achievement as well as the accomplishment of specific objectives embodied in the programs. Our strategy in this effort calls for the collection of student achievement data on at least two occasions, and for the collection of information about the nature and extent of program-related instruction the students received.

Procedurally, the implementation of the evaluations of the SEDL and SWRL programs shared some common characteristics. To the extent possible, evaluation data collection was coordinated with the existing structure set up by the labs for the installation and maintenance of their programs. In particular, both programs depended on a key individual at each site, the "district coordinator" for liaison work. The evaluation for each of the programs also worked through these district coordinators in obtaining counts of participating teachers and students, securing cooperation in data collection, and in monitoring the return of data from the sites. The CSE evaluation also took advantage of the pre- and inservice training sessions scheduled by the labs, both to observe these activities and, on the same field visit, to observe the use of the programs in classrooms in the vicinity of the regional training sessions.

Program-Specific Elements in the Design of the SEDL and SWRL Evaluation

As highlighted in the previous chapter, the SEDL and SWRL approaches to developing and implementing BSLC programs were quite different. These differences reflect both the intent to serve somewhat different populations as well as more general differences between SEDL and SWRL in the manner in which

they interpret their mission as regional R & D laboratories. One key difference between the labs concern the extent to which ongoing service efforts, such as the BSLC program, are viewed as instrumental for obtaining information about product improvement. Specifically, SWRL operations resources kits were designed through a method of systematic product development whereby data related to product effectiveness and improvement was routinely gathered throughout the development process and incorporated into subsequent product revisions. The object of this process is the development of materials that "stand alone", that is, that can be used effectively by various teacher-groups with a wide range of experience. This feature of the SWRL program provides justification for the relatively reduced emphasis placed in the evaluation on detailed observation of SWRL-BSLC training and inservice activities, incomparison to the level of field effort devoted to the corresponding SEDL program components. The SEDL approach to product development follows more closely an "inservice" model where materials are a small part of the delivery system for the program which also includes teacher orientation activities and lab instruction of participating teachers in the use of the materials.

A second feature of SWRL's approach to the BSLC program (and to their programmatic R & D work in general) that influenced evaluation design decisions was their collection of "operations resources data" in routine program operations. From one point of view, these data gathering activities generate information useful in the prospective inquiry process. From a second perspective, they constitute mechanisms for ongoing self-monitoring and improvement of BSLC program products and services. Finally, SWRL's BSLC data collection activities (e.g., the Learning Center Accomplishment Summary and Student Placement Aid instruments) are in part focused on

issues of concern to CSE's evaluation of the program. While the need for objectivity and independence in the evaluation has been the primary factor in the development of CSE's data collection strategies, issues of duplication of effort and minimization of respondent burden were given equal consideration. For example, economies in both these regards were realized in the collection of SWRL-BSLC implementation and student assessment data; the details of these arrangements are discussed in the sections to follow.

Another important way in which the evaluation strategies for the two labs differed follows from the subject matter embodied in their materials; SWRL's program is monolingual English, whereas SEDL's is Spanish/English bilingual. The major impact of this language difference on the evaluation was in the area of the construction and selection of appropriate achievement measures for students in the bilingual program. Our approach to student assessment for the SEDL program acknowledged the need to provide a range of testing options so as to accommodate both the potential diversity in participating students' language skills as well as the concerns of critics of the standard practices in the evaluation of bilingual programs. In addition, the bilingual nature of the SEDL program created a need in the evaluation for sensitivity to the mix, in implementation, of Spanish and English instructional components. Finally, the SEDL program places relatively greater emphasis on personnel intensive training and technical assistance activities, so the indices of implementation and perceived quality of the program directly incorporate information about these program features.

Instrumentation

SEDL-BSLC IMPLEMENTATION SURVEY

Our plan for the evaluation of the SEDL-BSLC program calls for the documentation of BOLaR and BCPM implementation for two basic reasons. First, it is essential to determine the extent and nature of the SEDL Learning Centers that were created as a result of the BSLC initiative, essentially to establish that the intended service was being provided. Second, the planned strategy of analyzing student outcomes called for indices of program implementation and quality to be related to variations in student achievement.

In each year of SEDL program operation; CSE constructed a teacher implementation survey to obtain information relevant to these two objectives; copies of each of these are contained in the Appendix. In constructing the first year survey, CSE staff generated a list of questions that would provide a description of how the BSLC materials were being used; which program components were being used; which students in terms of English or Spanish limited speakers or bilinguals are using the BSLC's, the amount of time spent on basic skills instruction; the criteria used for placement into the program and the quality of the BSLC training. Items were developed for the SEDL Implementation survey and were reviewed by CSE senior staff members. The survey was then based on their input and issued in final draft form at an SEDL inservice training session midway through the first program year. At the same time the survey was sent to SEDL for comment and subsequently revised before its administration in spring, 1978.

The survey instrument developed for use in the second year of SEDL program operations drew heavily on the first year survey, but also included

coverage of several additional concerns. The plans for the content of the new survey were presented to representatives of CSA, SEDL and NIE during meetings held at UCLA in January, 1979. Discussions at these meetings led to the identification of several areas in need of clarification, as well as a few new issues of interest. The major revisions in the second year instrument focused on probes into the amount of BSLC instruction and its distribution across language of instruction; student groups, and modes of instruction; additionally, several terms used in the survey were defined therein in order to eliminate variability in the interpretations respondents attacked to them. The final draft of the revised instrument was sent to CSE's bilingual review panel and to SEDL, NIE and CSA, and was approved for in data collection.

MEASUREMENT OF SEDL PROGRAM IMPLEMENTATION

Objective-Based Tests

Objectives-based tests were developed for the SEDL component of the evaluation. They were designed to assess the extent to which students achieved the objectives stressed in the BOLaR and BCPM programs. In constructing measures that would be instructionally sensitive to program effects, CSE staff completed a two step process. The first step required that the major objectives taught in the BOLaR and BCPM programs be identified. Secondly, CSE staff reviewed and selected measures that would accurately assess the objectives that had been identified.

CSE staff identified the major skills taught in the BOLaR and BCPM programs that were also representative of standard reading and mathematics objectives at the second and third grade levels.

The BOLaR materials, (both Spanish and English reading) and BCPM materials were reviewed, and the skills in which most instruction and practice were provided were identified as appropriate objectives. Objectives which reflected transfer skills in mathematics, such as word-problems and double digit addition and subtraction in order to ensure representativeness of on grade level (2 and 3) content area skills, were also chosen.

The objectives were reviewed by content area experts, including those with expertise in bilingual education and sent to SEDL for their comments and review.

CSE planned to locate test items that would adequately assess each of the objectives, and instruments which would meet CSE's financial constraints. The number of objectives identified was constrained by the need to limit student test time, while assessing each objective with an adequate number of items.

CSE staff reviewed the resources that were available at the center. Projects such as the Early Childhood Education Evaluation (ECE), the Evaluation of Reading in the Los Angeles Unified School District - Area B, the Bilingual Prediction Study and the Systems for Objectives-Based Evaluation in Reading (SOBER/Spanish - SOBAR/English) had developed, validated and administered criterion-referenced tests to assess student achievement as part of their evaluation. CSE also had access to the CSE Elementary School Test Evaluation as well as other test reviews funded through NIE support units and the state of California, for the Early Childhood Evaluation study.

Of particular interest to CSE staff however was the Systems for Objectives Based Evaluation in English (SOBAR) and Spanish reading (SOBER). These criterion-referenced tests were developed from a comprehensive set of objectives designed to measure student skills in encoding, decoding, vocabulary and read-

ing comprehension. For each objective selected, three items were included to assess the objective. The directions, item content, vocabulary, illustrations, format and procedures were excellent for use with Mexican-American students in grades kindergarden through third. The User's Guide for SOBER/SOBAR suggested that the tests could be used for pre- and posttesting as well as a diagnostic or prescriptive purposes.

As one of the criteria for selecting objectives was to give consideration to general grade level based skills that are part of the standard school curriculum, the system that the SOBER/SOBAR tests offered facilitated the process for the development of CSE objective-based tests. The objectives and items that the system offered in Spanish and English was congruent with the BOLaR objectives that CSE staff had identified. The SOBER/SOBAR items were easily adaptable for CSE's test construction purposes in terms of formating and leveling of vocabulary in both languages. Also among the pool of possible items to be used in the evaluation were those developed specifically for the BOLaR program (see CSE Quarterly report for the period April 1, 1977 to June 30, 1977 for more details on these early tests).

Items constructed to assess student achievement in mathematics for the BCPM program were drawn from various sources. Test items that had been constructed under the direction of Dr. Walker adequately assessed the objectives current CSE staff deemed important for the BCPM program. CSE staff however modified much of the Spanish to suit the grade level of the students being tested. CSE staff also utilized story problem items from the ECE Evaluation and translated them into Spanish. The story problems were designed to be administered orally, in both Spanish and English, to the student so as not to be confounded by language or reading ability.

The English reading tests, BEA and BEB were created to measure skills taught in the BOLaR programs. Students were assessed on the following objectives:

1. Given a word orally that has a single initial consonant and given a set of written words; the learner will select the written word with the same initial consonant.

2. Given a word orally that has an initial consonant blend and given a set of written words, the learner will select the written word with the same initial consonant blend.

3. Given a word orally, the learner will demonstrate sight vocabulary recognition by selecting it from a set of written words.

4. The learner will demonstrate word attack skills by selecting from among written alternatives an orally presented regularly spelled word.

5. Given a set of written sentences with the verb missing and a set of written inflected forms of a verb, the learner will select the form that correctly completes the sentence.

6. Given a written selection and a written question about one of the details of the selection, the learner will select the sentence that best answers the question.

Each of these tests contained 30 items.

The Spanish reading tests SpA and SpB consisted of five subtest areas. Students were required to achieve the following objectives:

1. Given a word orally that has a single initial consonant, and given a group of consonants, the learner will select from among written alternatives the initial consonant of the word.

2. Given a word orally and given a group of vowels the learner will select the double vowels in the word.

3. Given an incomplete written sentence and a set of written words, the learner will select the word whose meaning is appropriate to the context of the sentence.

4. Given a word orally and a group of written words, the learner will demonstrate vocabulary recognition by selecting the written form of the word.

5. Given a written selection and a literal comprehension question about the selection, the learner will select from among written alternatives the answer to the question.

Each of these test contained 26 items, 4 fewer than the English reading tests.

The Math tests, MTA and MTB covered eleven subtest areas corresponding with the major skills taught in the BCPM program. Students were required to achieve the following objectives:

1. Given a number written in the form of "x" tens and "y" ones, the student will select the corresponding two-digit numeral.

2. Given a two digit numeral, the student will choose a representation of the number as "x" tens and "y" ones.

3. Given two or three single-digit addends in horizontal format, the student will select the sum from among four alternatives. (sums through 19; no carrying)

4. Given two double digit addends, the student will select the sum from among four alternatives. (no carrying)

5. Given a subtraction problem in horizontal format, the student will select the difference from among four alternatives. (subtrahends to 19; no borrowing)

6. Given a subtraction problem with a double digit subtrahend and a double digit minuend, the student will select the difference from among four alternatives. (no borrowing)

7. Given an equation of the form $a + ___ = b$, $___ + a = b$, or $a - ___ = b$, the student will choose the missing number from among four alternatives.

8. Given a pictured set of objects the student will select from among four alternatives the number of sets of objects that are contained in the pictured set.

9. Given a set of coins whose total value is less than 99¢, the student will select from among four alternatives the total value of the coins.

10. Given a pictured clock face, the student will select its standard time notation from among four alternatives.

11. Given a story in both English and Spanish that describes an addition or subtraction problem, the student will select from among four alternatives the number sentence that corresponds to the problem.

Each of these tests contained a total of 24 items.

As part of the test development process, CSE staff sent the objectives-based tests to SEDL and the bilingual advisory panel for their comments and review. As well, CSE senior staff, particularly those with expertise in bilingual test construction in the content areas of reading and math reviewed the instruments. No substantive changes were recommended.

After the first administration of the tests in April, 1978, CSE staff conducted an item analysis. After examining students' responses, modifications were made. In order to make the difficulty level more appropriate for the grade level of the student population being served, item stimulus and/or the response distracters were changed accordingly.

In summary, items finally selected to assess students' skills accomplishment with the BOLaR and BCPM programs were chosen to be consistent with the objectives identified. Two parallel forms of each test were thus constructed using the following item sources:

BOLaR/Spanish — SOBER/Espanol, Science Research Associates
Bilingual Prediction Study, CSE

BOLaR/English — SOBER/English, Science Research Associates
Early Childhood Education Evaluation, CSE

BCPM — PACT test, State of Wisconsin
Primary Mathematics Behavioral Objectives and Test Items, Downer's Grove, Illinois
Mathematics Management System, Clark County, Nevada
Early Childhood Education Evaluation, CSE

Norm-Referenced Achievement Measure

In response to the central mandate for the BBLE evaluation, CSE administered a norm-referenced measure to students participating in the SEDL program in order to procure data comparable with national performance standards.

The issue of testing ethnolinguistic minority students to measure their achievement levels in reading and mathematics is a complex one. The problem becomes not whether to test them, but rather how to do it in a manner that accurately assesses their specific abilities, and in a manner that does not create a bias either against them or in favor of them.

The logic and structure of a test must be commensurate with the logic and structure of a student's own linguistic and cultural background. Achievement tests, for example, are often linguistically biased in favor of students who are proficient in English and/or whose native language is English. Mathematics concepts as often test verbal and reading skills in English thereby putting the non-English reading student at a distinct disadvantage.

In addition, psycholinguistic research indicates that language structures reflecting a child's level of cognitive development often appear earlier in one language than they do in the other for children learning two languages. Research also indicates that test scores have been found to rise when students are tested in their native language.

The evaluation of bilingual instructional programs must ensure that adequate procedures for assessing a given child's educational status are utilized. It is important to assess the students' level of language proficiency in English and the native language, as not to confound achievement with lack of English language skills. Evaluations of bilingual programs

cannot take for granted that all students, though participating in bilingual instructional programs, are receiving the same amount of instruction in English and Spanish in the content areas of reading and mathematics, for example, and/or are equally proficient in both languages. Thus, conclusions drawn regarding student outcomes will be different depending on the language in which performance is measured.

The SEDL program is bilingual and makes provision for reading and mathematics instruction in both English and Spanish. CSE staff wished to employ both Spanish and English assessment measures so as not to limit documentation of program effects. By using a standardized measure as well as an objectives-based measure, CSE staff planned to be able to separate those influences which would have been operating had the SEDL program not been implemented from program related influences on performance, respectively.

In the identification of standardized instruments to be used in the SEDL component of the evaluation, consideration was given to a broad range of tests. Guidance for the initial review of instruments was provided by *a Guide for Assessment Instruments for Limited English* Speakers (Pletcher, B.; Locke, B.; Reynolds, D.; Sisson, B. New York: Santillana Publishing Company, 1978). This catalogue provides information on assessment instruments related to the following criteria:

- Designed for use directly with students in grades K-6
- Used currently in the United States
- Normed with students residing in the United States
- Developed, adapted, or statistically tabled for use with students whose first language is Chinese, Portuguese, French, Italian, Navajo, Tagalog, or Spanish
- Designed to yield scores which can be interpreted outside the context of a specific curriculum or commercial series

- Considered by the author or source to be complete technically and with regard to instrument content
- Not designed to identify developmental or physiological problems
- Available to the public

The initial review of instruments for use with the SEDL program also involved the application of criteria for instrument assessment developed by CSE (Hoepfner, R. CSE Elementary School Test Evaluations, 1970). The CSE criteria examined instruments for measurement validity, examinee appropriateness, and administrative usability. CSE project staff further reviewed the tests for linguistic and cultural relevance.

Among the measures considered for review for English language and reading and mathematics were the following:

- California Test of Basic Skills (CTBS), Level C, Form S
- California Achievement Test (CAT), Level II
- Boehm Test of Basic Concepts, Level K-2, Form B
- Metropolitan Readiness Test, Level II
- Stanford Early Achievement Test, Level II
- Inter-American Test of Reading and Numbers, Level 3-4
- Tests of General Ability, Level II

Among the measures CSE staff considered for review for Spanish language and reading and mathematics were the following:

- California Test of Basic Skills (CTBS) Español, Level C, Form S
- Prueba de Habilidad General--Inter Americana, Level II
- Short Test of Educational Ability--Spanish Edition
- Stanford Early Achievement Test--Chicano Spanish Version, Level II
- Prueba Boehm de Conceptos Basicos, Form B

The CSE test review and selection process involved two stages. First, CSE staff reviewed the descriptions provided in The Guide for Assessment

Instruments, thereby limiting the scope of the exercise. Second, CSE requested from the appropriate publishers specifications of the following instruments:

- CTBS English--CTBS Español, Level C, Form S
- Test of General Ability--Prueba de Habilidad General, Level II
- Stanford Early School Achievement Test (SESAT)--SESAT Chicano Spanish Version, Level II

The decision made by CSE to review only these three instruments was based on the fact that versions of the same test were available in both English and Spanish. In addition, CSE staff requested any technical information that was available and information describing the test development process, in particular for the Spanish versions of the tests.

At the same time, CSE staff reviewed the SEDL program objectives, the scope and sequence of the skills assessed in English and Spanish reading and mathematics, the general format of the materials, and the method which students were required to use in responding to the exercises provided.

Since CSE staff was still very concerned about testing students in the appropriate language based on their level of language proficiency in English and Spanish, it was viewed as desirable to employ a language proficiency measure to individual students. This posed an important problem however in terms of time and financial constraints, because such assessment typically requires students to be individually tested by the teachers or aides.

Another alternative was to base the proficiency decision on teacher judgment. Pursuing this alternative, however, was not feasible since it would depend on the extent individual teachers were cognizant of bilingual education goals, and the type of social, cultural, and the educational training/orientation

the teacher had toward bilingual instruction and ethnolinguistic minority students. Again, due to time constraints, in particular, CSE staff opted not to utilize this procedure.

Upon receipt of the specifications from the test publishers, CSE staff reviewed them with regard to the SEDL program objectives, appropriate norming samples, and the cultural and linguistic relevance of the test items and illustrations. The Comprehensive Test of Basic Skills (CTBS), Level C, Form S, was found to most accurately reflect the objectives of the SEDL-BSLC curriculum. In addition, CTB/McGraw-Hill provided CSE staff with useful information regarding the CTBS test development process for the English and Spanish versions. Of particular interest were the following points:

- The CTBS/English Version was standardized with a national sample which included 7.9% Spanish speaking students.
- The U.S. Office of Education sponsored bias studies on the English reading, language and mathematics sections of the CTBS. A panel including Asian, Mexican-American and Indian educators evaluated items which appeared to be statistically biased against certain ethnic groups and concluded that only 91 items out of 933 were potentially biased against Spanish-speaking students.
- A background questionnaire was completed for each student involved in the standardization study.
- The CTBS/Español was adapted from the CTBS/English. Comparable scores in both languages could be generated. Both language versions have statistically similar levels of difficulty, reliability and validity.
- The adaptation of the CTBS/Español is based on CTB/McGraw-Hill's efforts to procure data based on students' reactions to vocabulary, reading passages and comprehension questions and mathematics concepts and applications.
- Vocabulary on the CTBS/Español was selected to match the familiarity of words comprising the English test.

Based on this information, CSE staff made the decision to use the CTBS, English and Spanish versions, to measure on grade level achievement for students participating in the SEDL program.

The Spanish/English Reading and Vocabulary Screening (SERVS), a pre-measure to determine whether a given student should be tested with the CTBS in English or Spanish was developed by the California Test Bureau (Monterey, California) to accompany the CTBS. The SERVS is designed to be administered in both English and Spanish to allow comparison of each student's performance in two languages. The SERVS provides practice in marking test responses, measures of oral comprehension of test directions and terminology and sentence comprehension items.

CSE staff, at a meeting in Austin, Texas at SEDL, announced their decision to use the CTBS instrument with SEDL staff members. CSE staff shared copies of the CTBS/Level C with them and presented a rationale for its use. SEDL staff members were satisfied with this decision

MEASUREMENT OF SWRL PROGRAM IMPLEMENTATION

Two sources of information about the installation and use of SWRL-BSLCs were considered in the evaluation. First, SWRL, as part of its regular program operations constructed and administered the Learning Center Accomplishment Summary (LCAS), which probes a variety of areas of concern to the evaluation. The attractiveness of relying on the SWRL data collection operation for implementation information was enhanced by the open collaboration between SWRL and CSE in the LCAS development process. The LCAS, although in slightly different forms, was administered to all participating SWRL sites during both years of program operation.

The second source of implementation information on the SWRL program comes from a set of questions contained on a cover sheet accompanying the SWRL End-of-Year student assessment instrument. Questions to be answered by BSLC teachers concerning individual students use of and success with the program are included on the cover sheet. A subset of these questions were also attached, as a cover sheet, to the CSE-developed criterion-referenced

test administered to SWRL students in the spring of the second year of program operations (1978-1979).

Objectives Based Tests

CSE staff agreed to use SWRL-developed instruments to measure student achievement with the SWRL program. There are two such program-developed measures specific to the reading and mathematics kits, the End-of-Year tests and the Placement Aid. These tests were administered at local site discretion as part of SWRL's ongoing project activities during the 1977-78 school year.

The End-of-Kit tests are objective referenced and designed so that students may check their progress on the skills in which they have received direct practice. The SWRL-developed Placement Aid is designed to assess a wide range of reading and mathematics skills to determine the appropriate level of placement into the SWRL kits. Items on the placement aid were drawn from the End-of-Kit tests. The End-of-Year test, which was used by SWRL as the measure to assess cognitive outcomes of the BSLC curriculum, is the Placement Aid, renamed. CSE staff view the SWRL End-of-Year test/Placement Aid as measures that can be used both for diagnosis, precription, placement, and for pre- and posttesting and mastery testing. The End-of-Year test and Placement Aid are useful as summative evaluation tools as they incorporate a wide range of objectives and difficulty levels standard to the kindergarten through sixth school curriculum. Accordingly, the student assessment information incorporated into the CSE evaluation of the SWRL program consists exclusively of the Placement Aid/End-of-Year test. This decision is consistent with agreements reached at the winter, 1978 meetings to limit first year student testing in criterion-referenced measures.

In October, 1978, CSE staff met with SWRL staff to discuss data collection plans at the SWRL sites for the 1978-79 school year. It was agreed that an objectives-based test comprised of items sampled from the SWRL placement aids

would be used as part of the evaluation to measure the attainment of program skills.

CSE staff, with permission from SWRL, chose to use the Placement Aid because data were available from previous administration of the instrument. Before CSE staff could actually formalize an objective-based measure from the Placement Aid prerequisite tasks were in order.

The Placement Aid was developed as a diagnostic tool and not designed as an instrument that would attend to individual differences. Students were not required to take the same sections of the test, nor were they given corresponding sections on the two occasions it was administered during the first year. CSE performed a preliminary item analysis to reduce the Placement Aid by eliminating items found to be inappropriate for students during past administrations. Item selection was based on data supplied by SWRL concerning past student performance on the Placement Aid when it was used for the purpose of Placement and as a posttest. CSE staff computed the percentage of students who had answered Placement Aid items correctly and specified the given grade levels at which the test had been administered, for each of the two test administrations. The levels of item performance were plotted on graphs which reflected grade level and test section.

In screening the Placement Aid items CSE staff developed specific criteria upon which to judge items for inclusion in the objectives-based measure, the criteria were as follows:

a. The difficulty of an item should decrease in response to two factors, exposure to schooling and sensitivity to appropriate instruction. That is to say, students at higher grade levels should find the items easier than students at lower grade levels. CSE staff further postulated that students should

perform better on the posttesting of the items as it would reflect use of the materials over time.

b. The items should represent a wide range of difficulty levels. Given that the ultimate task will be to interpret program effects, if all students fail particular items after having used the materials, or reversely pass particular items before having received basic skills instruction, these kind of data will not aid in describing student outcomes with regard to BSLC instruction.

c. Only sections of the test pertaining to the SWRL materials at the second and third grade levels will be considered for inclusion in the CSE item pool.

It should be noted that without a detailed analysis of the instructional experience of the students responding to the items, these performance comparisons provide relatively rough information about the item characteristics.

On the basis of these criteria, items were classified into three groups:

a. Completely acceptable--No objection to the item, it should be used if coverage on the corresponding curricular objective is desirable.

b. Provisionally acceptable--Use the item only if it is needed to achieve adequate coverage on the corresponding curricular objective.

c. Unacceptable--The item is not intended for the appropriate grade range. It could also be the case that the item, though appropriate for grades two and three, will not yield meaningful information about student outcomes.

The CSE item selection process identified 40 reading items and 19 mathematics items as being potentially sensitive to the effects of BSLC instruction with the SWRL materials. It is important to note, however, the item pools for both reading and mathematics did not reflect all sections of the Placement Aids. The reading item pool consisted of all 76 items in the first 9 sections

of the Reading Placement Aid; while the mathematics item pool consisted of 36 items in the first 8 sections of the Mathematics Placement Aid. Both the reading and mathematics item pool corresponded to curriculum use in the second and third grades.

CSE staff identified a subset of items for inclusion in the CSE objectives-based measure. There were a total of 37 SWRL Reading Placement Aid items, 5 reading comprehension questions and 33 SWRL Mathematics Placement Aid items. It should be noted that the reading passages and comprehension questions used on the CSE instrument for SWRL were originally developed for the SEDL English Reading test. CSE staff chose not to use the reading comprehension items in the SWRL Reading Placement Aid because they contained vocabulary and sentence structures too difficult for second and third grade. It was felt, however, that since reading comprehension is an important skill to assess, the CSE objective-based instrument should attend to the issue.

SWRL Norm-Referenced Achievement Measures

As was the case with SEDL, CSE was bound by contractual agreement with CSA to provide SWRL student achievement data interpretable in relation to national performance standards. In agreement with SWRL staff, CSE chose sub-tests in reading and mathematics from a norm-referenced instrument in order to measure the attainment of second and third grade level skills for students participating in the SWRL program.

CSE has developed[1] a "test match procedure" for locating a test appropriate

---

[1]Through the efforts of Elizabeth Weinberger and Clint Walker on the Criterion-Referenced Test Evaluation Project and of Lynn Morris in writing the Program Evaluation Kit.

for the evaluation of a particular instructional program. The test used for assessing the achievement of objectives of an instructional program should be one that allows the evaluations to base their selection on relevance. The test match procedure developed by Morris and Fitz-Gibbon is a method whereby the relative congruence of a specific program's instructional objectives and the items on a standardized test can be determined.

The first step in the procedure required CSE staff to identify the instructional objectives for the SWRL program. To accomplish this task CSE staff examined and reviewed the objectives provided for the teachers using the basic skills program, the SWRL specifications used in developing the program, and the instructional materials.

CSE staff found the SWRL objectives to reflect accurately the content of the program. However, the objectives were presented to the teachers in the form of scope and sequence charts and did not specify the stimulus conditions under which the student would be asked to demonstrate skills mastery. In order to establish the behavioral aspects of the objectives, CSE staff consulted the SWRL specification used in developing the program. This process facilitated CSE staff in also determining whether the format of the various test items was consistant with the format of the student practice exercises.

It is important to note that the test selection process was modified to accomodate CSE's evaluation interests with regard to the SWRL program. Morris and Fitz-Gibbon recommended in their book How to Measure Achievement: Evaluation Kit (Los Angeles, California; Sage Publications, 1978) categorizing the objecitves with respect to their emphasis in the program's implementation. CSE staff however, felt that the available SWRL data regarding program implementation, at that point in time, was not extensive enough to warrent making

such decisions, and opted instead to weigh the objectives in accordance with the SWRL specifications.

After identifying the program objectives CSE staff with the assistance of Ms. Gayle Straus, (a doctoral student in Learning and Instruction at UCLA), selected three standardized tests to analyse: the Comprehensive Test of Basic Skills, (CTBS) levels B and C; the Stanford Achievement Test, (SAT) primary level, and the California Achievement Test, (CAT) levels 11 and 12. These six test forms were examined for the match of their items with those portions of the SWRL materials most likely to be used by students in grades 2 and 3. CSE staff determined that students at these grade levels would be working with the following kits:

Basic Reading Skills I - Yellow Kit
Decoding Skills I - Yellow Kit
Basic Mathematics Skills I - Purple Kit
Basic Mathematics Skills II - Yellow Kit
Basic Mathematics Skills III - Red Kit

The tests were chosen based on the information presented in CSE's publication, Elementary School Test Evaluation Series. The CSE evaluation provided information on the technical adequacy of the tests as well as information regarding the nonmative data established by the tests publishers.

Having identified the SWRL program objectives and three tests to examine, CSE staff followed the procedure to construct a test comparison chart. The test comparison chart displayed the following information: the number of different objectives which the SWRL program addressed; the SWRL objectives that matched the test items on the CTBS, (levels B and C); the CAT (level 12); and the SAT (primary level 1); a rating regarding the importance of each objective; the closeness of fit between item content, format and instructional materials; and the appropriateness of difficulty of each test item.

The ratings assigned to each test item were used to compute summary figures and indices for each of the standardized tests under consideration. The Grand Average described the average, per test item, of the combined ratings of importance, curricular match, and item difficulty. It produced a summary figure for determining how well, overall, a test covered important program objectives. CSE staff felt however that the actual numerical values assigned were not as important as the comparison of values for the tests under consideration. The Index of Coverage showed how completely each test covers the SWRL program objectives. It is based on the proportion of program objectives that a test measures. Unlike the Grand Average, this index does not assess the quality of the items, only whether or not they correspond to the program objectives. Its possible values range from a low of 0 to a high of 1.0. The data presented in the chart of test comparison shows that three of the four test actually measured fewer than half the skills taught in the BSLC program.

The Index of Relevance showed which proportion of the test was comprised of items which assessed skills other than those addressed in the SWRL program. This index was calculated by dividing the number of items on the test which corresponded to SWRL program objectives by the total number of items on each of the norm-referenced tests. As in the Index of Coverage, the possible values could range from 0 to 1. The results, as they appear on the chart, showed that for the CTBS, CAT and SAT, the Index of Relevance was higher than the Index of Coverage, ranging from the CTBS (level B) which had no relevant items to the SAT in which 24% of the items were irrelevant.

The CSE test match procedure produced the finding that the CAT, level 12, Form C was the most appropriate to the SWRL materials. As is displayed in the Chart of Test Comparison, the CAT showed the highest Grand Average of the

Table 1

Test Characteristics Computed via the Table for Program-Test Comparison

Test and Level	Appropriate Grade of Students	Grand Average	Index of Coverage	Index of Relevance
California Achievement Test (CAT) level 12	1.5 to 2.9	1.61	0.43	0.93
California Test of Basic Skills (CTBS) level C	1.6 to 2.9	1.41	0.49	0.85
Stanford Achievement Test primary level 1	1.5 to 2.9	1.33	0.71	0.76
California Test of Basic Skills (CTBS) level B	K.6 to 1.9	1.94	0.46	1.00

three tests.

The Chart of Test Comparison displays an Index of Coverage and an Index of Relevance for the tests being examined. At first glance it appears in the chart that the Index of Coverage is low for the CTBS, SAT and CAT. The chart also shows however that the CAT, level 12, Form C to have a 0.93 Index of Relevance. This high index means that although the CAT might not adequately cover the SWRL program in its entirety, it does comprise items which do assess skills being emphasized by the Program. CSE's staff decision was to use the CAT, level 12, Form C. This decision was based on the following criteria:

- congruency with SWRL objectives
- on-grade level of assessed students
- limited likelihood of producing ceiling effects.

It is important to note that the formal test match procedure was carried out for the SWRL reading materials in particular. CSE staff did however informally compare the CAT mathematics items with the SWRL mathematics objectives to determine if the difficulty level, format and skills being assessed by the CAT were congruent with the SWRL mathematics materials.

Data Collection and Sampling

The data collection plan for both SWRL and SEDL was developed to gain information regarding the following questions:

- Are the benefits of the program being extended to the proper target group(s)?
- Are the SWRL and SEDL materials being used to the extent and in the manner to which they were intended?
- Are the intended student outcomes being achieved?
- Can relationships be found among student and teacher characteristics and program features that bear on student outcomes?

SEDL-BSLC IMPLEMENTATION SURVEY

1977-1978 SEDL-BSLC Implementation Survey

The sampling plan for the SEDL Implementation Survey included all teachers, aides and other school personnel involved with the SEDL program. CSE staff contacted SEDL designated coordinators regarding the first wave of data collection in April, 1978. The intent of employing the services of the district coordinators was that SEDL staff had previously established solid channels of communication at each of the 60 sites.

CSE staff phoned individual coordinators in order to explain their position as third party evaluators; to outline data collection needs; and to solicit their cooperation in collecting the data. CSE staff requested district coordinators to distribute the surveys to all teachers and aides within their districts, and to also coordinate the return of the surveys to CSE. To facilitate the process each district coordinator was asked to provide to CSE staff the following information:

- estimates of the number of BOLaR and BCPM teachers and/or aides at each of the participating schools in their district,
- estimates of the number of BOLaR and BCPM students at each grade level, at each of the participating schools

- the name and address of a BSLC contact person at each school

In addition, district coordinators, in districts which had also been selected to participate in CSE's testing effort for the 1977-78 school year were informed of the testing plan and its purpose.

On May 5, 1978, CSE staff mailed survey packets to district coordinators with instructions for distributing the packets. CSE staff instructed coordinators to distribute the packets to all teachers and aides participating in the BSLC program within their district. A cover letter to individual coordinators explained the purpose of the survey, emphasizing the importance of collecting implementation data in relation to the BSLC program and the need to obtain a high return rate on the surveys.

Packets distributed to each teacher and aide included the survey; a letter describing the BSLC evaluation project and a postage paid envelope for returning the survey to CSE. Approximately 942 surveys were mailed to 60 district coordinators for distribution.

Although all district coordinators had agreed to participate in the evaluation effort as per our initial telephone contacts, four coordinators did not distribute the materials. Most of these coordinators felt that having to fill out the survey would cause teachers undue pressure so close to the end of the school year. One coordinator in particular refused to give the teachers in that district the option of filling them out becuase she believed it was not within their contract. Also, during this period of time one district to whom CSE staff had sent surveys, had dropped the BSLC program.

By the end of June, 1978, 269 surveys had been returned to CSE, representing 75% of the districts participating in the SEDL Basic Skills program. Follow-up telephone contacts were made to those districts which had not returned any surveys.

1978-1979 SEDL Implementation Survey

In the fall, 1978, CSE staff contacted district coordinators at the SEDL sites regarding the objectives-based and standardized tests that would be administered as part of the data collection effort during the 1978-79 school year. At that time, CSE staff inquired as to the number of teachers and aides within each of the districts. In March, 1978, CSE staff sent letters to each of the district coordinators who would be participating in the second wave of data collection with the survey instrument.

The process of data collection was identical to that of the spring, 1978. CSE staff requested that district coordinators distribute the survey to all teachers and aides involved with the SEDL Basic Skills program and coordinate the return of the surveys to CSE. A total of 760 surveys were administered to 65 teachers and 40 aides within 12 districts in Arizona; 98 teachers and 88 aides within 18 districts in California; 89 teachers and 74 aides within 7 districts in Colorado; 105 teachers and 32 aides within 6 districts in New Mexico; 113 teachers and 59 aides within 13 districts in Texas. (See the Appendix for a detailed chart.)

As of August, 1979, a total of 179 surveys had been returned representing 24% of the number of surveys sent. The total sample represented 12 schools in Arizona, 8 schools in California, 16 schools in Colorado, 16 schools in New Mexico, and 13 schools in Texas. (See Appendix for a more detailed description.)

SEDL-BSLC TESTING INSTRUMENTS

Objectives-Based Measures (1977-78)

The CSE developed objectives-based instruments designed to assess Spanish reading, English reading and mathematics in relation to the SEDL Basic Skills program were administered for the first wave of data collection in the spring

of 1978. The purpose of collecting data at the end of the first year of program implementation was so that CSE staff would be able to provide baseline data regarding students' attainment of skills addressed in the BOLaR and BCPM components.

The sampling plan focused on 24 school districts within all five states being served by the SEDL Basic Skills program: Arizona, California, Colorado, New Mexico and Texas. The CSE instruments sampled students in second and third grade participating in the SEDL Basic Skills program. CSE staff based their decision to sample students at those grade levels on previous telephone contacts to SEDL coordinators. These contacts revealed that teachers using the BOLaR and BCPM programs most heavily focused BOLaR and BCPM instruction with second and third grade students.

The strategy for data collection at each of the SEDL sites was to rely on the district coordinators to distribute the materials to teachers and to also coordinate the returns of tests to CSE. CSE staff contacted district coordinators by phone in May, 1978, to inform them of the testing that would take place in their districts and to solicit their cooperation. CSE staff informed coordinators of the following information:

- The objectives-based tests would reflect assessment of specific reading and mathematics skills addressed in the SEDL program.
- Each test would require approximately 45 minutes of student time.
- The test would be group administered by the classroom teacher or aide in the appropriate language.
- Tests were to be administered to second and third grade students participating in the BOLaR and/or BCPM programs.
- Teachers had the option of administering the tests in English and/or Spanish, reflecting the most proficient language of the student.

In addition, CSE staff inquired as to the numbers of second and third grade students using the BOLaR and BCPM program components in each of the schools within their districts; and the closing date of school.

On May 10, 1978 packages containing reading and/or mathematics tests were mailed to district coordinators for distribution to teachers and aides, who in turn administered the tests. The test packets included the reading and/or mathematics test; a cover letter to teachers explaining the purpose of the tests; general directions for test administration; specific instruction for each of the tests; a roster for listing the students being tested with spaces provided for the teacher/aide to indicate BCPM program of instruction and language of program instruction on an individual student level; and postage-paid mailers for returning completed tests. In total, 1778 reading tests, and 1197 mathematics tests were mailed for administration, corresponding to the numbers obtained from district coordinators.

Approximately one week prior to the closing date of each school district, CSE staff contacted district coordinators to determine the status of the spring 1978 testing effort. CSE staff inquired as to whether the correct numbers and types of tests had been received, whether testing had been completed or was still in progress and whether any problems had arisen regarding the overall testing effort. In addition, CSE staff informed district coordinators at that time that as part of the Basic Skills evaluation, CSE would be administering revised versions of the objectives-based measures, and would also be administering standardized achievement measures, pre- and post, during the 1978-79 school year.

By July, 1978, CSE staff had completed data tracking and processing. In total 688 English reading tests, 812 Spanish reading tests, and 896 mathematics tests were received, representing a return from 100% of the districts

sampled, and 76% of the schools to who tests were sent. (See the Appendix for a table describing actual numbers of tests sent and received from each district.)

SEDL Objectives-Based and Standardized Testing (1978-79)

For the second wave of data collection, CSE staff administered two types of assessment measures of student performance, pre- and post. Students were assessed with both a revised version of the objectives-based tests that had been administered during year one of program implementation (1977-78), and the California Test of Basic Skills (CTBS), level C, Form S, a standardized achievement measure designed to assess students' reading and mathematics skills. CSE staff used the standardized measure to compute student outcomes with regard to national performance standards.

It was of interest to CSE staff to retest at sites that had participated in the first wave of data collection during the 1977-78 school year. Again testing efforts would be focused on second and third grade students participating in the SEDL program.

CSE employed the same strategy in collecting student outcome data during the second year of SEDL program implementation. In October, 1978, CSE staff contacted SEDL district coordinators to inform them again of the testing that would take place, and to inquire as to the approximate numbers of students who would receive reading and mathematics tests. At the same time, CSE staff informed district coordinators that in addition to teachers having to administer standardized achievement tests in English and/or Spanish, they would first be asked to administer a language indicator called the Spanish-English Reading and Vocabulary Screening (SERVS) designed to assess the language in which given students were to be tested with the CTBS. Each district coordinator was responsible for obtaining from teachers within their districts

the number of CTBS needed in Spanish and in English, based on the outcome of the language screening measure.

In mid-October, 1978, CSE staff packaged and sent objectives-based tests and SERVS to 26 school districts in all 5 states that were participating in the SEDL Basic Skills programs. Within a three week period of time, CSE staff contacted district coordinators to determine the status of the testing effort and to obtain the actual numbers of CTBS needed in English and Spanish to administer to students. Upon receipt of this information, CSE staff packaged and mailed CTBS tests to district coordinators at each of the sites. In turn, coordinators distributed the CTBS to teachers within their districts. Approximately one month later, CSE staff contacted district coordinators to determine the status of the testing effort with the CTBS. At the same time, CSE staff informed coordinators that posttesting would occur during April, 1979.

By January, 1979, all fall data had been tracked. A total of 1467 objectives-based English reading tests, 910 objectives-based Spanish reading tests, and 799 bilingual mathematics tests had been administered and returned to CSE. As for the standardized achievement measures (CTBS), a total of 814 CTBS-English and 348 CTBS-Español had been administered and returned to CSE. (See the Appendix for a more detailed account of these numbers in relation to specific schools within districts.)

In March, 1979, CSE staff contacted SEDL district coordinators to inform them of the posttesting that would occur in their districts in April. Based on the numbers of objectives-based tests returned from the fall, 1978 data collection effort in Spanish and English, CSE staff packaged and sent both objectives-based tests and CTBS to district coordinators to distribute to teachers in their districts.

The testing sample included all 26 districts that had participated in the fall, 1978, with the exception of West Las Vegas, New Mexico. It is perhaps important to note that this site was of particular interest to CSE staff because equal numbers of students (205) had been tested in both English and Spanish with the objectives-based tests in 4 out of 6 schools; and equal numbers of students had also been tested in both languages in 2 out of 6 schools with the CTBS. This indicated to CSE staff that a majority of the student population were potentially equally proficient in Spanish and English.

CSE staff felt that the West Las Vegas site was a substantial loss. The reasons, as West Las Vegas school personnel explained, were politically based. Of the entire school administration, those who viewed bilingual instruction as a viable alternative in education were to be relieved of their responsibilities within the district. CSE staff contacted SEDL staff immediately to inform them of this problem. Another major concern additionally at this point was the implementation and use of the SEDL materials. Due to the change that had taken place within the district at the administration level, it was realistic to believe that "special programs" implemented for bilingual instructional purposes may be functioning at a minimum. Both CSE and SEDL staff were in close contact by phone and letter with the West Las Vegas superintendent of schools, the assistant superintendent, and the person who was assuming the role of Director of Instructional/Special Programs and the SEDL/BSLC district coordinator regarding these matters. By February, 1979, it was apparent that the West Las Vegas site would not participate in the CSE evaluation.

By mid-June, 1979, objectives-based test data and CTBS data had been returned to CSE. A total of 875 objectives-based English reading tests, 463 objectives-based Spanish reading tests, 729 objectives-based mathematics

tests, 456 CTBS-English tests and 102 CTBS-Español tests had been administered and returned. This represented an 83% return from the 24 districts sampled, and 76% of the schools to whom tests were sent. (See Appendix for a table describing actual number of tests sent and received from each district for spring, 1979).

SWRL-BSLC TESTING INSTRUMENTS

1977-78 Data Collection

Data collection efforts at SWRL sites for the 1977-78 school year involved a cooperative arrangement between CSE and SWRL. This agreement entailed the allocation of funds to SWRL which CSE would have spent collecting data on the SWRL program. CSE agreed to support processing of a subsample of the first administration Placement Aid test data and the printing, administration, and processing of the second administration of this instrument.

Data collection during the first year of program implementation focused on the fall and spring administrations of the SWRL Placement Aids and End-of-Year tests, respectively. It was felt that the SWRL-developed measure would adequately reflect the evaluation needs and interests of CSE staff, since the first program year focused on the implementation and use of BSLCs. CSE staff felt that the SWRL Placement Aid would lend itself to making informative statements about program outcomes.

The sampling plan for the first year of SWRL program implementation included the school districts which SWRL had identified for more intensive data collection. The sample was comprised of 26 districts in 19 states, with a total of 93 participating schools. The states in which SWRL-BSLC programs had been implemented were: Arizona, California, Colorado, Idaho, Kansas, Montana, Nebraska, Nevada, New Mexico, Minnesota, North Dakota, Oklahoma, Oregon, South Dakota, Texas, Utah, Washington, and Wyoming.

SWRL Objectives-Based and Norm-Referenced Testing (1978-79)

Data collection efforts for the second year of program implementation involved two assessment measures: an objectives-based test composed of items sampled from the SWRL developed Placement Aids, and reading and mathematics subsets from the California Achievement Test (CAT), a norm-referenced measure.

The sample was comprised of 17 school districts which SWRL had designated as part of their intensive sample, i.e., a subset of the 26 districts referred to above. On the basis of analysis, CSE performed on the SWRL Placement Aids from the 1977-78 school year, a decision was made to concentrate student data collection efforts at the second and third grade levels in each of the districts.

In cooperation with CSE, SWRL staff composed a letter to be sent to district coordinators in each of the 17 districts describing the data collection effort as part of CSE's evaluation activities for SWRL. CSE staff requested permission from SWRL staff to contact the SWRL district coordinators by telephone to appraise them of the testing that was to take place in April, 1979.

CSE staff obtained an updated list of coordinators names, addresses, and phone numbers from SWRL staff. From March 10-26, 1979, CSE staff contacted SWRL district coordinators. Since CSE's evaluation efforts were to focus on second and third grade, fourteen sites were eligible to participate in the second wave of data collection given this condition. To facilitate the process, CSE staff relied on the district coordinators to distribute the tests to teachers and/or aides within their districts and to coordinate the return of the testing materials to CSE. SWRL district coordinators were advised by CSE staff of the following information:

- Both the objectives-based and CAT would arrive on or about March 17, 1979.

- Testing should take place in each of the schools during the second and third weeks of April in grades 2 and 3 only.
- Testing should be done on two consecutive days--the objectives-based measure on the first day and the CAT on the following day.
- Both the objectives-based measure and the CAT would take approximately 1½ hours to administer.
- CSE staff would send coordinators a prototype for a memo that was to be circulated to teachers before the tests arrived at the districts.

In addition, CSE staff inquired as to the numbers of second and third grade students at individual schools within each of the districts being sampled. CSE staff assured SWRL district coordinators that they would be available should any questions, concerns, or problems arise.

CSE staff ordered the appropriate numbers of CATs as designated by district coordinators. The CATs were ordered from the California Test Bureau, Monterey, California, in late February, 1979; however, data collection was slowed down considerably because the tests did not arrive at CSE until April 5, 1979.

Upon receipt of the CATs, CSE staff packaged and mailed packets to district coordinators at each of the 14 sites. The packets included a cover letter to coordinators; a letter to teachers explaining the purpose of the testing; the objectives-based measures with instructions for administration; and the CATs with an administrator's manual for each teacher, and postage paid return envelopes for each of the schools within each of the districts. A total of 1827 objectives-referenced tests, and 2195 CATs were packaged and mailed.

By mid-July, 1979, CSE staff had completed data tracking and processing. In total, 717 objectives-based tests and 1037 CATs had been administered and returned. This represented an 86% return of the districts sampled, and a

73% return from the schools to whom tests were sent. (See the Appendix for exact numbers regarding number of objectives-based and CAT's sent to individual schools within districts.)

FIELD VISITS

As part of the data collection effort CSE staff attended SEDL and SWRL inservice workshop during the 1977-78 and 178-79 school years. The reasons for attending these meetings were twofold. First, CSE staff wished to gather informal, qualitative data on program participants' reactions to the SEDL and SWRL programs. Secondly, CSE staff wished to observe the link between technical assistance in the form of on-site training by the laboratories and program operations which would be validated by on-site observation of program implementation at the classroom level. Inservice worshops generally consisted of the following for both SEDL and SWRL:

- implementation techniques
- guidelines for managing program components
- teachers' diagnosis of students level of skills accomplishment upon entering the basic skills program
- reporting requirements for participants

In addition, CSE staffs' on-site visits yielded information on the variability of use of SEDL and SWRL materials within schools at individual districts. Of the data collected, CSE staff found the following information to be of particular interest;

- The SEDL Spanish reading and oral language materials were being used as core bilingual program instructional material at many of the sites.
- The SWRL program was being used for skills reinforcement/remediation in grades 6-8 at many of the schools in New Mexico.

(See Appendix for detailed description of inservice workshops and on-site classroom observations at SEDL and SWRL sites in the form of field notes.)

CHAPTER IV: RESULTS

Results of the CSE evaluation of SEDL and SWRL-BSLC programs will be presented in two sections. The first section discusses results obtained from the SEDL evaluation: the Implementation survey, student test data, and relationships among program implementation variables and student achievement. Section two presents analogous results for the SWRL program; results related to program implementation, student tests, and finally a discussion of the relationships between implementation and outcome variables.

SEDL Results

Discussion of SEDL results is organized in the following manner: First, a description of the program derived from the Implementation Survey is presented. (Refer to Appendix A10.) This discussion is followed by a presentation of student test data, which includes a summary of the psychometric properties of program achievement measures and a description of student performance on each of these measures. The third part of this section presents the relationships between program implementation variables and student outcome data presented. The most important variables contributing to student achievement gains within the SEDL-BSLC program are isolated and the implications of these process variables for producing gains in achievement in future BSLC programs are discussed.

SEDL IMPLEMENTATION SURVEY

Description of Program Participants

The SEDL Implementation Survey was returned by 179 teachers in the spring (May) of 1979. Respondents were predominately full time (94%) classroom teachers or other certified school district employees, such as

resource teachers and project specialists (13%). Teaching experience ranged from one to forty-one years. The majority of the sample, however, had been teaching eight years or less (65%). Fifty-eight percent of the teachers earned a BA or BS degree and 35% held Master's degrees. The sample included one Ph.D. and a small number of instructors with only high school diplomas or certificates of completion (7%).

One hundred and fifty seven teachers (88%) reported that they had experience teaching in bilingual programs; of this group, 84% had five years or less bilingual teaching experience. Interestingly, the number of teachers reporting that they were bilingual-biliterate in English and Spanish was 66, which is only 37% of the entire sample. Thus, there is evidence that some of the teachers participating in bilingual programs were not bilingual-biliterate.

Teachers were sampled from school districts in five states - Arizona, California, Colorado, New Mexico and Texas. Three school districts participating in previous data collection efforts were absent: West Las Vegas (Nevada), Harrison (Colorado) and Pecos (New Mexico).

Description of Program Setting

SEDL materials were used primarily in the lower elementary grades; 96% of the sample taught in kindergarten to grade 3. Class sizes ranged from 1 to 99 students; however, 54% of the classrooms had from 24-30 pupils. The mean class size for BSLC instruction was somewhat smaller than regular class size, 21 as opposed to 26 pupils.

BSLC materials were most often housed in self-contained classrooms (77%) rather than in separate resource centers (14%). Students working in the materials tended to be grouped with others in their regular classrooms (82%); only 17% of the sample worked the students from different classes. A

majority of the teachers (72%) used BSLC materials in small group settings, 3-9 students. The remaining survey categories of instructional organization--whole class, large group, and individualized--were each used about half as often (29-35%) as the small group mode. Large group and whole class instructional modes were reported as never used by the largest percentages of respondents 22% and 30%, respectively

Assignment of Students to Program

The two most important placement criteria cited for assignment of students to BSLC programs were teacher evaluations (65%) and language dominance tests (43%). Placement tests were most often accorded moderate importance. Achievement tests had little weight as selection criteria; 50% never used standardized bilingual achievement tests, 46% never used standardized English tests, and 39% never used grades as selection criteria. I.Q. tests were uniformly rejected as a placement criterion with 71% of the sample reporting that such results were never used for assignment.

Although achievement tests and grades had little weight as selection criteria for a large number of the respondents, 54% of the teachers reported that BSLC students differed from non-BSLC students in achievement level. When asked to compare BSLC and non-BSLC participants, 40-49% of the respondents reported that BSLC students differed on the dimensions of language background, socioeconomic status, and previous school experience. A majority of the teachers (61%) reported that BSLC and NBSLC students were similar in "cultural background".

Survey results do not indicate which combinations of selection criteria were most often used to place students into programs or what factors

entered into the criterion "teacher evaluation." From these data, few conclusions can be drawn about the relationship between selection criteria and kinds of students selected for BSLC instruction. It would appear that a majority of the students in BSLC programs are not different demographically from those in non-BSLC programs, but that they do differ in achievement from their peers. If this is so, then students have indeed been placed according to the CSA mandate.

Instructional Goals

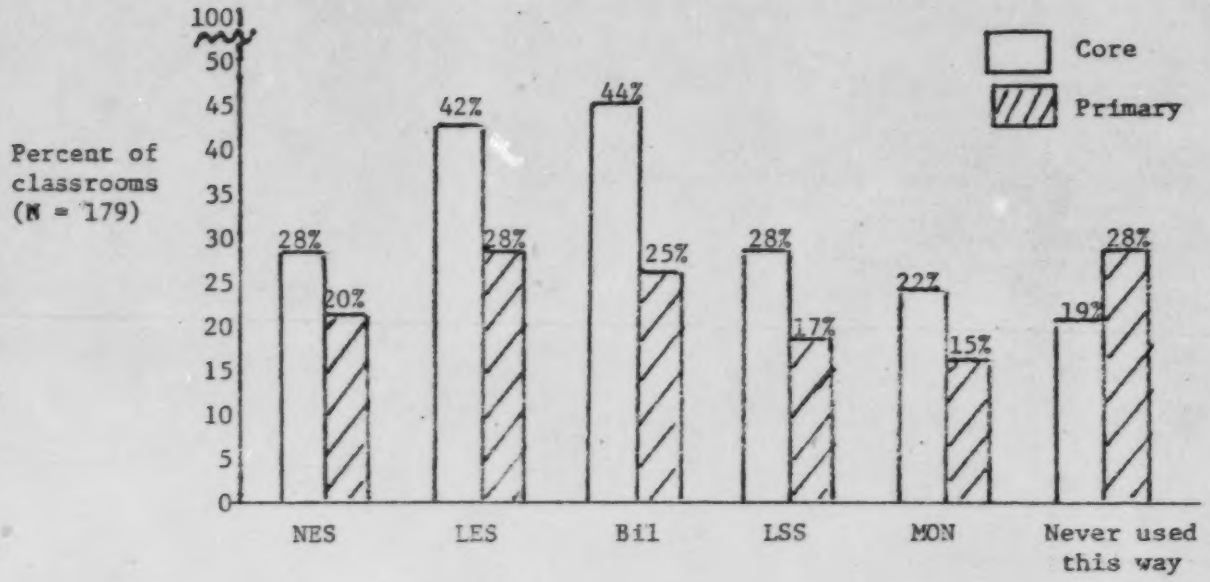
Respondents most often reported that their primary goal for BSLC students was to "provide enough competence in English and Spanish so that instruction in both languages is possible" (41%). There is some evidence that this goal was interpreted to mean improvement of the English skills of non-English speakers (NES), limited-English speakers (LES), and bilingual speakers, rather than the Spanish skills of limited-Spanish speakers (LSS) and English monolinguals. Thirty-one percent of the sample stated that they used the materials to facilitate a transition from Spanish to English instruction, while only 6% selected "strengthening and maintaining Spanish skill" as an important goal.

Use of Materials with Students

The CSA mandate for BSLC programs was to provide remedial instruction for rural students from low income families. Figure 2 displays the percentage of teachers reporting core vs. remedial instruction for the five language subgroups served by SEDL. Use of materials as core exceeds their use in remedial programs for all groups. Additionally, the number of teachers reporting that they never use the materials for remediation exceeds those stating that they never used them as primary instruction.

FIGURE 2

Percentage of Classrooms Using SEDL Materials as Core v. Remedial Instruction for Students with Diverse Language Skills



KEY: NES - Non-English Speaking
LES - Limited English Speaking
Bil - Bilingual Speakers - English and Spanish
LSS - Limited Spanish Speaking
MON - Monolingual English Speaking

Although it appears that BSLCs are most often used as core curriculum for all groups, the manner in which sample teachers report using BSLC materials suggests that the materials were used for different purposes with different language subgroups. English monolinguals most often were assigned BSLC materials as supplementary or enrichment activities (27%). By contrast, materials were used with bilingual students in a variety of ways: as core curriculum (44% of the classes), as supplementary materials (41%), and as enrichment (46%). Fully 37% of the respondents reported that they never used the materials for Spanish instruction with LSS and English monolingual students, while 55% reported that they did use it for English instruction with NES and LES students and ESL students (English as a second language). Only 17% reported assigning SEDL materials to English monolinguals as part of a bilingual curriculum; on the other hand, over one third of the teachers did use it with LES or NES students for bilingual instruction.

Respondents most frequently (44%) reported that the BSLC materials seemed most effective with bilingual students; others reported that the materials were most effective with LES students (37%), NES (25%), LSS (22%), and English monolinguals (18%). These perceptions support the results reported in the previous section related to program goals and assignment of students to materials. Whether teacher perception of the materials' effectiveness with particular language subgroups is a result of the kinds of students initially selected for BSLC instruction, e.g., predominately Spanish-dominant, or whether it is a result of how materials were used with certain subgroups cannot be determined from the survey responses.

Language of Instruction

The average number of minutes per day in BSLC instruction in English exceeded that in Spanish for all program components except oral language, where instructional time in each language was essentially equal. Instructional time spent in each language was more comparable within programs, BSLC or non-BSLC, than between. Within non-BSLC instruction, the ratio of mean instructional time in English to Spanish ranges from 2.5:1 for reading and oral language to 3:1 in mathematics. Mathematics instruction, however, was predominantly conducted in English in both BSLC and non-BSLC programs, reflecting perhaps the nature of the skill being taught, or the language dominance of those responsible for presenting instruction and explaining mathematical concepts.

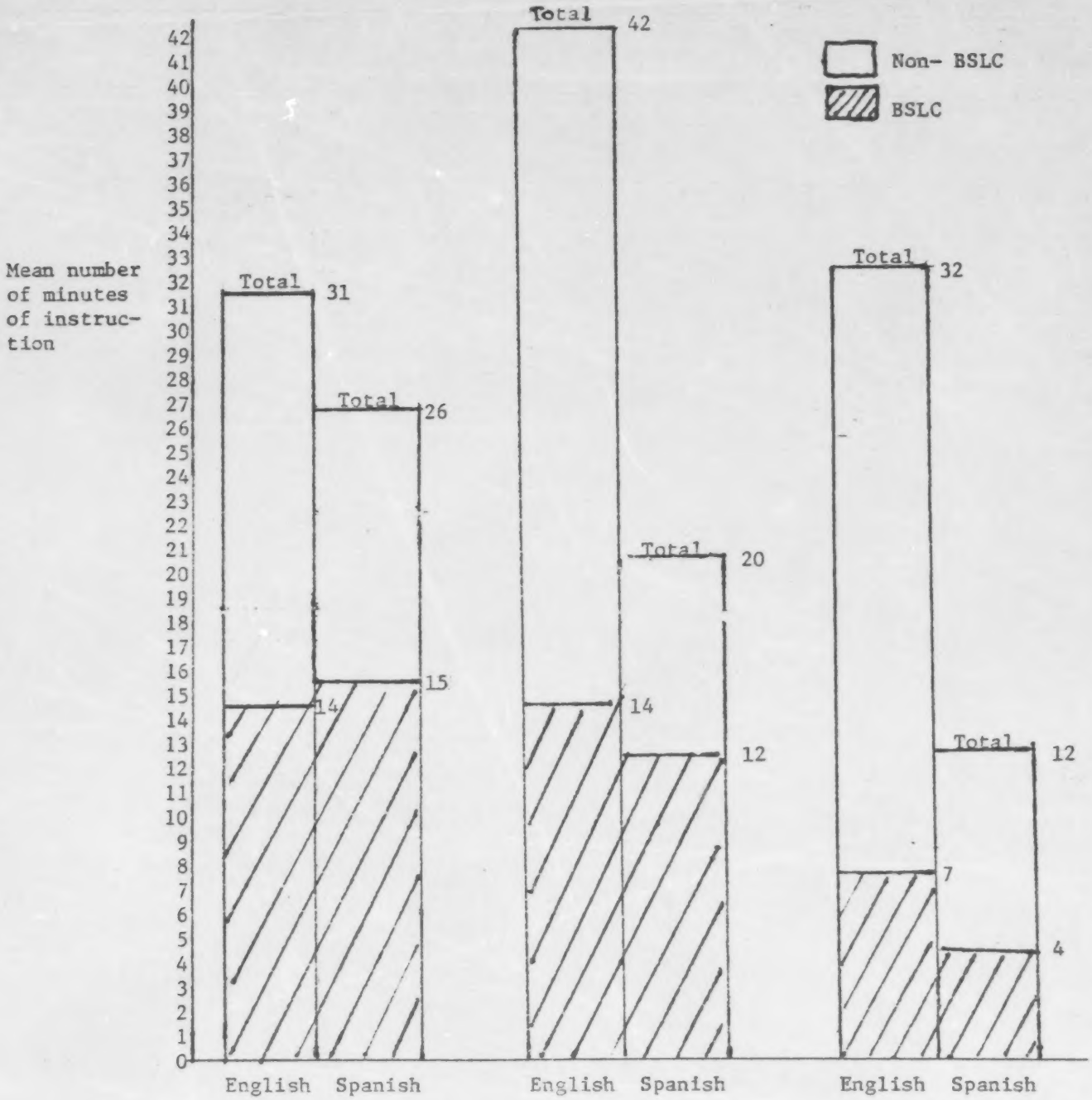
Figure 3 displays the relationship of English to Spanish instructional time for BSLC and non-BSLC programs. Although, with the exception of oral language, non-BSLC instructional time exceeds BSLC instructional time, a greater proportion of instruction is conducted in Spanish in BSLC than in non-BSLC. The greater emphasis on Spanish in BSLC programs as compared to non-BSLC instruction may have much to do with the BSLC materials themselves. That is, the language of instruction in any program component may be a function of the availability of appropriate Spanish or English teaching materials.

Coverage of SEDL Program Materials

Data from the Pupil Program Information Forms (PPIF) administered by SEDL were used to construct two indices of student progress through the SEDL materials. Using classroom level summaries of PPIF information, containing percentages of students completing subunits within the materials, we first formed a variable reflecting the amount of material covered by

FIGURE 3

Mean number of minutes per day spent in SEDL/BSLC v. Non-BSLC Basic Skills Instruction in English and Spanish



certain percentages of the class. The variable changes value each time 20% of the students in the class complete a unit in the SEDL curriculum or 100% of the students are working in a subunit. For example, if the PPIF summary for a given class indicates that 100% of the students in the class completed only one unit in the materials, the index would be 5. An index of 5 could also mean that 20% of the class are working in 5 different units. If, for another class, the PPIF showed 60% of the class completing two subunits, that class would be assigned a value of six on the amount of coverage variable.

The second variable constructed from the PPIF summary simply counted the number of subunits within a level of the curriculum (BOLaR has four levels, and BCPM has two) that any students in a given class completed; this variable is referred to as breadth of coverage.

Results for the amount of coverage variable are displayed in Figure 4, for each level of the BOLaR and BCPM programs, in Spanish and English. The bars show the relative amounts of coverage for the two programs between the fall to spring (1978-1979). For each program, there are substantial increases in the amount of material covered. The figure indicates that within BOLaR there are slightly greater increases made in the English component than Spanish, except for level II. Between BOLaR and BCPM, greater increases are observed in the latter.

Figure 5 shows the corresponding breakdown for the breadth of coverage variable and suggests a similar progress differential favoring the BCPM program.

Teacher Attitude Toward Program

Five questions on the SEDL Implementation Survey summarized respondents' attitude toward the program (questions 14, 15, 18, 19 and 21). These items required teachers to rate separately the quality of BSLC materials, training

FIGURE 4

Amount of Coverage of SEDL Program Materials

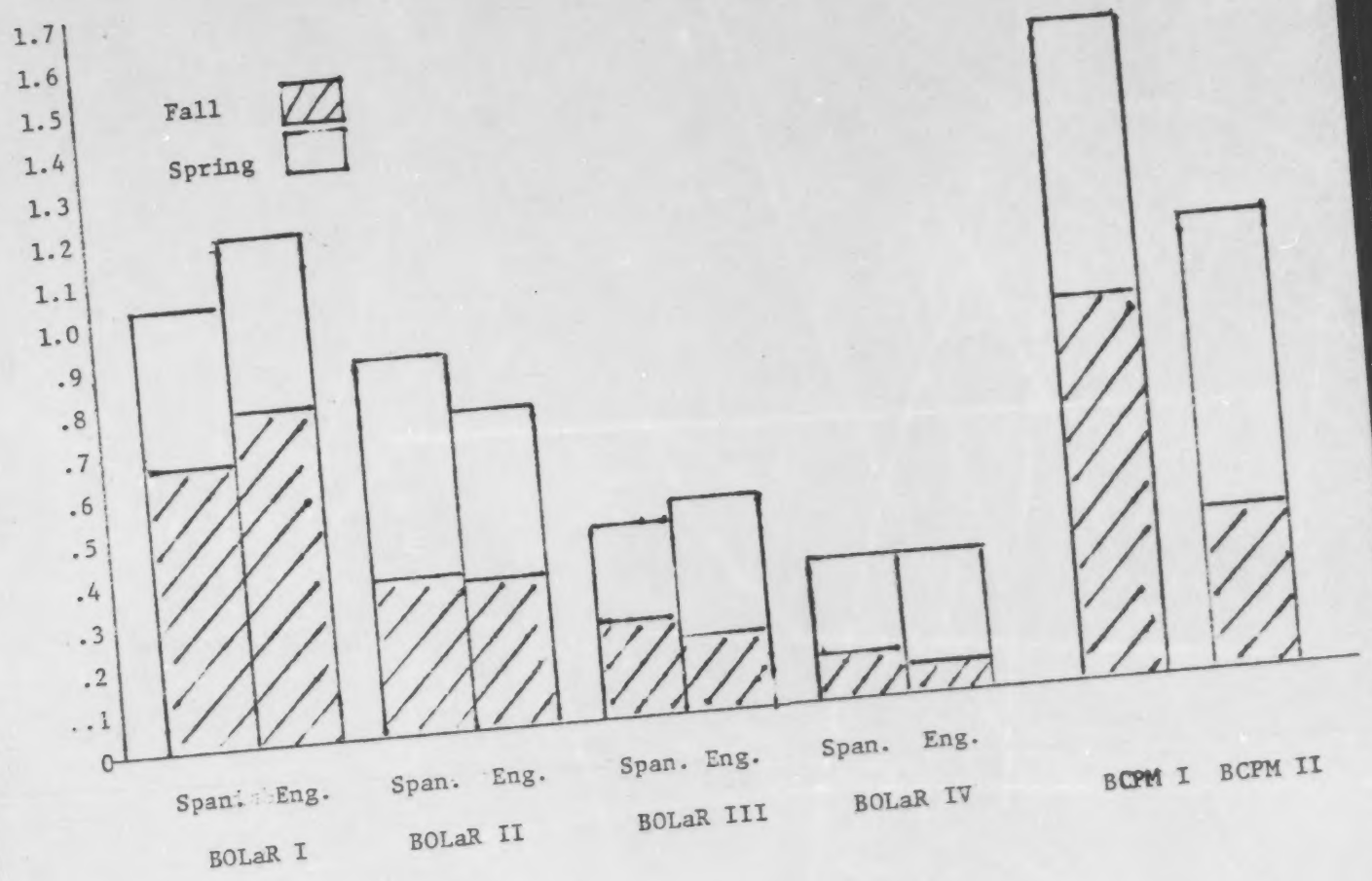
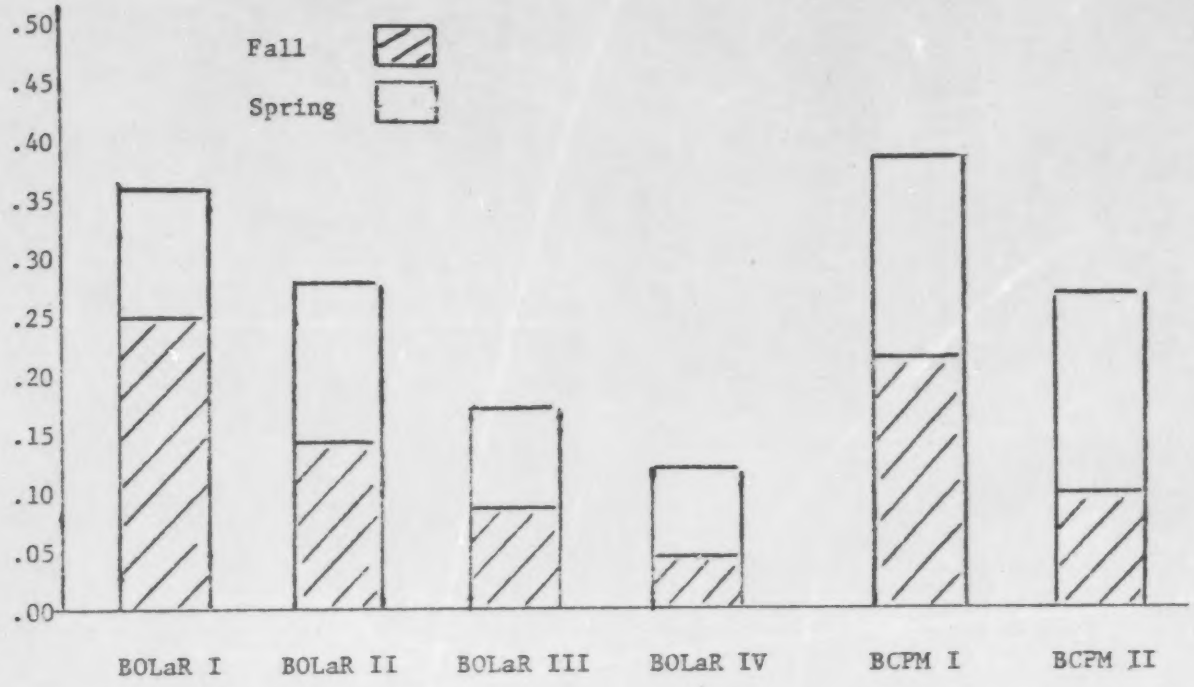


FIGURE 5

Breadth of Coverage of SEDL Program Materials



and technical assistance along a variety of dimensions. Separate indices were computed to characterize attitude toward each of these three program components. In addition, respondents were required to rate the extent to which the use of SSLC materials was producing gains in student achievement as well as their overall reaction to the program.

Figure 6 displays the attitude toward each of the program components (materials, training and technical assistance), compared with overall attitude toward program and perceived gains in student achievement. When the percentages of teachers responding at the two extremes (i.e., good vs. poor) are compared, responses were quite positive on all dimensions. In general, respondents are quite positive in their overall reaction to the program. They are least positive in their perceptions of the extent to which the program is producing educational gains for students. To illustrate, 69% of the respondents reacted quite favorably to the program, but only 48% answered that they definitely thought it was producing student gains. In viewing teacher perceptions of the three program components, responses were about technical assistance were most positive, followed respectively by training and materials. Note, however, that relatively few respondents rated the quality of technical assistance. A majority of respondents (64%) indicated that they had not required technical assistance which can be interpreted as additional evidence for positive attitude toward SEDL training and materials.

Table 2 displays intercorrelations among selected attitude indices (see Appendix for complete item-intercorrelations). This table reveals some interesting relationships. One would expect a strong and positive relationship between question 15, the degree to which teachers think the program is bringing about educational gains for students, and question 21, overall quality. Paradoxically, a larger percentage of teachers favor the program than feel

FIGURE 6

Teacher Attitude Toward SEDL's BSLC Program: Materials, Training, Technical Assistance Compared with Attitude Toward Overall Quality and Perceived Program Effectiveness

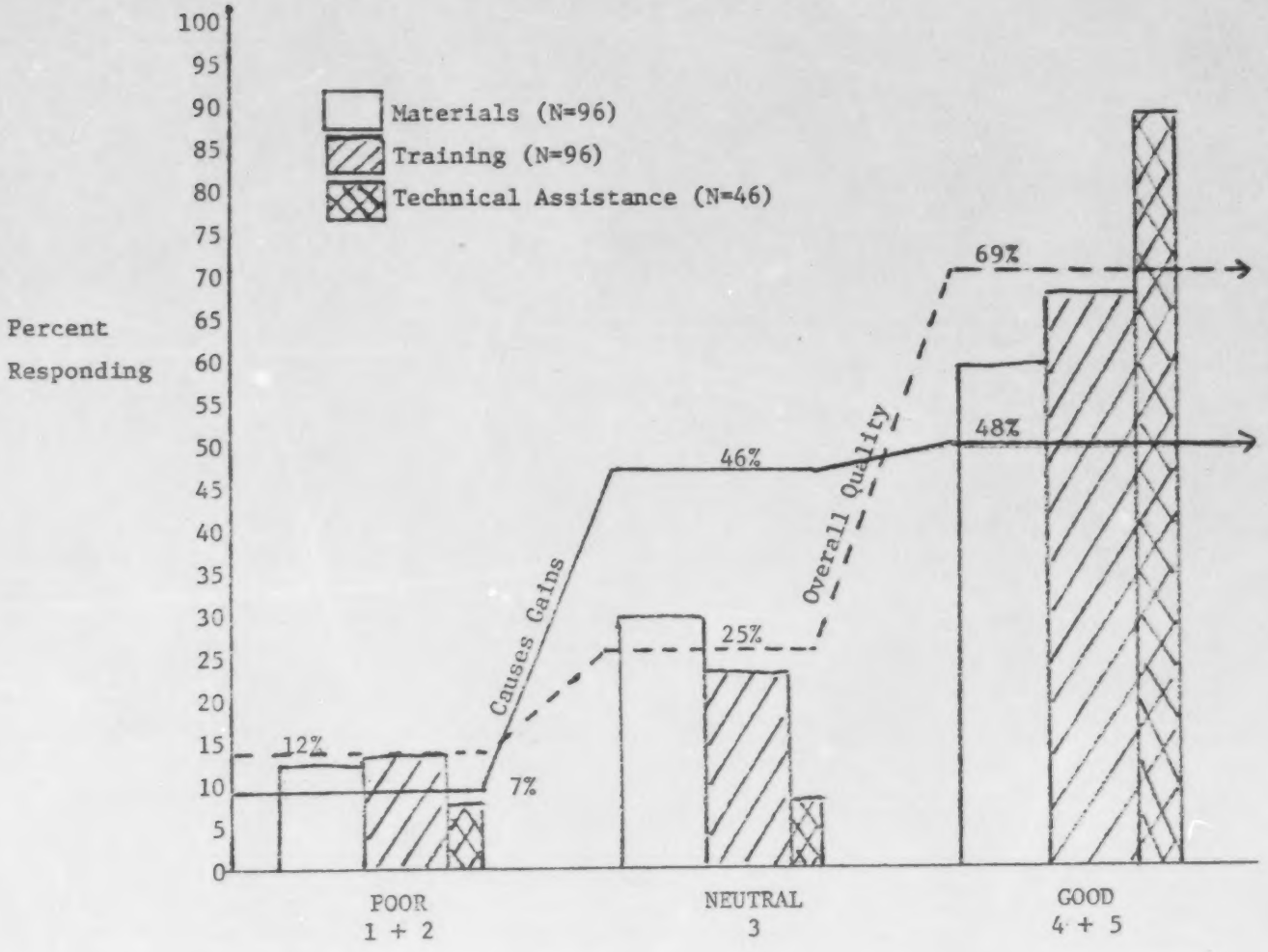


TABLE 2

Intercorrelations of SEDL Implementation Teacher Survey Items Measuring Attitude Toward Program

	14b	15	18 a-c	21
14b Materials				
15 - Perceived Student Gains	.497			
18 a-c Training	.417	.238		
21 Overall	.613	.392	.531	

the program is having an effect on basic skill attainment. In addition, survey item-intercorrelations indicate that perceived overall program quality is associated more strongly with quality of materials and training than with perceived student gains. Thus, respondent definitions of "good programs" are not limited to the production of instructional gains; quality of materials and of the instructional support systems are also important qualitative dimensions.

SEDL STUDENT TEST DATA

Psychometric Qualities of Outcome Measures

Seven measures of program outcomes were administered to BSLC pupils on two test occasions - fall, 1978 and spring, 1979. Three of the tests were developed by CSE staff to measure program-specific outcomes: a Spanish reading criterion-referenced test (SP-CRT), an English reading criterion-referenced test (E-CRT), and a mathematics criterion-referenced test (M-CRT). Test content was specified prior to item construction and approved by SEDL staff; CSE then developed test items congruent with content specifications to match the curricular goals of SEDL materials. The test development process is described more fully in instrumentation.

In order to enhance the generalizability of evaluation results beyond program-specific objectives, four norm-referenced achievement measures were selected by CSE staff. The CTBS Spanish Reading and Mathematics (CTB-SR and CTB-SM) and CTBS English Reading and Mathematics (CTB-ER and CTB-EM) were chosen as measures of SEDL program goals through a test-matching procedure described previously. Thus two sorts of interpretations about student performance could be made: Data from the CRTs would provide information about student achievement vis-a-vis program objectives, and NRT results would allow comparisons between SEDL participants and a national norm group.

Several sets of test statistics were calculated for the seven student outcome measures--SP-CRT, ECRT, MCRT, CTB-SR, CTB-SM, CTB-ER, and CTB-EM--used in the evaluation. Item difficulty indices (p. values) alpha coefficients, means and standard deviations were obtained for each of the subscales of the seven measures. These statistics are reproduced in the Appendix. The alpha coefficients for the Spanish Reading CRTs ranged from .36 for the lowest subscale retained in the data analysis to .84. Alphas for the English Reading CRT ranged from .21 to .80 Mathematics CRT reliabilities ranged from .28 to .81 CTBS Spanish Reading and Mathematics subscales ranged from .87 to .93, while the CTBS English Reading and Mathematics test scales ranged from an alpha of .82 for the lowest scale to .95 for the highest. These ranges include all test subscales on both the fall and spring test administrations.

Although seven dependent measures were administered during the evaluation, not all seven were used in the data analyses. The two CTBS Spanish tests, Reading and Mathematics, were deleted from the regression analysis due to small sample size (n=59).

Student performance. Table 3 displays the pre- and posttest performance of the analysis sample on all seven program measures. Significant gains were obtained by the sample on all tests. In the absence of a comparison group or baseline data, it is difficult to determine how much of the gain is directly attributable to the program and how much may be due to non-program related factors (e.g., regression effects, maturation, and pre- posttest reactivity). The data do provide two valuable kinds of information for future implementation efforts. In the absence of tests to measure program-specific SEDL objectives, the Spanish, English, and Mathematics CRTs were developed. Results from these tests can provide baseline information against which future comparisons may be

TABLE 3

Group Gains in Reading and Mathematics from Fall 1978 to Spring 1979: SEDL

TEST		Fall	(N)	Spring	(N)	Net Gain	T value T_d	df
Spanish CRT Reading (N items= 26)	$\bar{x}$	16.92	904	19.14	516	2.22	8.53*	1418
	S.D.	4.98		4.44				
English CRT Reading (N items=30)	$\bar{x}$	19.40	522	21.29	889	1.89	5.10*	1409
	S.D.	7.35		6.63				
Math CRT (N items=34)	$\bar{x}$	23.74	823	26.77	731	3.03	9.20*	1552
	S.D.	6.99		6.08				
CTBS-English Reading (N items=153)	$\bar{x}$	95.48	311	109.45	311	13.98	12.72*	310
	S.D.	36.98		33.17				
CTBS-English Math (N items=53)	$\bar{x}$	31.49	311	35.12	311	3.63	5.12*	310
	S.D.	12.10		15.57				
CTBS-Spanish Reading (N items=74)	$\bar{x}$	34.65	60	40.90	60	6.25	4.71*	59
	S.D.	13.61		14.92				
CTBS-Spanish Math (N items=47)	$\bar{x}$	30.35	60	36.38	60	6.03	6.47*	59
	S.D.	9.43		9.36				

made if the program is continued. Similarly, no comparative data for rural, low SES, bilingual students exists for the CTBS tests. Our results may indicate to some extent how like groups of students may be expected to perform on these measures.

Some statements about the student achievement do follow from the results obtained on the CTBS English Reading and Mathematics tests. Results on these two measures by grade and by subtest are displayed in Table 4. Raw scores on the pre- and posttest have been converted into percentile rankings which reflect the status of the SEDL sample vis-a-vis the national norm group. As the CTBS was administered in December, pretest rankings were obtained from the December-January-February tables in the CTBS technical manual (CTB/McGraw-Hill, 1974). Pre- and posttest raw scores were converted by referring to the April-May-June tables.

Student CTBS scores were also converted into Normal Curve Equivalent scores (NCE). NCEs are linearly transformed scores with a mean of 50 and a standard deviation of 21.06 (see Tallmadge and Wood, 1976). The advantage that NCE scores have over percentiles is that the distances between NCEs are equivalent, making it possible to add or subtract the scores in order to examine student gains. Examination of the sample NCE scores reveals that students gained vis-a-vis the national sample in two areas, reading vocabulary and spelling; they maintained their position in mathematics computation. While the evaluation design does not allow us to attribute the gains entirely to program effects, they suggest possible areas of SEDL program strength for investigation in future studies.

TABLE 4

Percentile Rankings of CTBS English Tests

for SEDL Students

Results for Total Battery (Reading and Mathematics)

	Raw Score		Percentile Equivalent			
	Pre	Post	Pre	NCE	Post	NCE
Grade 2	110.7 (434)	127.9 (258)	38%	44	34%	41
Grade 3	160.0 (335)	164.3 (171)	66%	59	56%	53

Results for CTBS Subtests

	Raw Score		Percentile Equivalent			
	Pre	Post	Pre	NCE	Post	NCE
Reading						
Vocabulary	21.4 (311)	24.4 (311)	39%	44	40%	45
Sentences	14.1 (311)	16.5 (311)	46%	48	40%	45
Passages	10.8 (311)	12.5 (311)	44%	47	37%	43
Language						
Expression	14.0 (311)	16.0 (311)	44%	47	39%	44
Spelling	24.3 (311)	28.0 (311)	39%	44	41%	45
Mechanics	10.9 (311)	12.1 (311)	59%	55	57%	54
Mathematics						
Computation	16.2 (311)	18.8 (311)	53%	52	53%	52
Concepts	15.3 (311)	16.3 (311)	45%	47	40%	45

Relationship of Implementation Variables to Student Achievement

Selection of implementation variables. CSE staff identified six variables embedded in the SEDL program that should have a significant influence on student achievement. The variables were divided into two categories, control variables, or those variables related to characteristics of program participants, and program variables, variables associated with the BSLC program itself. The three control variables are student background, as represented by pretest scores and grade level, and teacher background, years of bilingual teaching experience.

The variable selected to represent teachers was the number of years of experience of bilingual teaching. It was thought that this predictor would contribute more to student gains than the simple index, years of teaching experience or years of education for two reasons. First, in the majority of evaluation studies, the variable "teacher experience" has no relationship to student outcomes (Lyon, et al.), and one of the characteristics unique to the SEDL remediation program was the bilingual component. Second, the relationship between years of bilingual teaching experience and bilingual reading competence should be positive. In the absence of data to indicate that no relationship exists between bilingual experience and student achievement gains, years bilingual teaching experience was felt to be a good predictor of student achievement. Two demographic variables related to the student sample were the grade level of the student and the pretest score. The pretest score served as an index of "instructional history," i.e., the background each student brought to the instructional setting.

One variable was selected to represent an overall index of the quality of implementation of the SEDL program, i.e., how well teachers actually

conformed to the SEDL conception of how the program should operate and how well teachers like the SEDL program. This index is a summary of several of the survey items which measure teacher conformity to implementation guidelines and teacher attitude and which were intercorrelated positively and significantly. Items 14a-g, 15, 16, 17, 18a-c, 20a, 20c, and 21 from the Implementation Survey were used in the calculation of the index of implementation quality. This index represents the standardized sum of responses to each item; the larger the sum, the more positive the quality of implementation, reflecting both adherence to program guidelines and attitude toward program. Intercorrelations among survey items comprising the index of implementation quality are displayed in the Appendix.

Two predictor variables were selected, one to represent the relative amounts of time spent in bilingual BSLC instruction and the other to represent total time of BSLC instruction, to reflect unique aspects of the SEDL-BSLCs. The bilinguality of the program was characterized by the ratio of English to Spanish instructional time within BSLC, summarized from survey question 5. Time allotted to remediation or use of SEDL materials was represented by the index, Total Time in BSLC instruction, also summarized from Implementation Survey Question 5. The control and program variables are summarized in Table 5.

Description of Implementation Variables: Descriptive Statistics and Intercorrelations

Table 6 displays the descriptive statistics for the program variables and five outcome measures, the English and Spanish reading and mathematics CRTs and the CTBS English reading and mathematics test. The analysis sample was predominantly second grade, although those students taking the norm-referenced and the Spanish reading CRT were somewhat older than students given the English reading and mathematics CRT tests.

TABLE 5

Description of SEDL Implementation Variables

Control Variables

Pretest	Mean pretest score of students on appropriate dependent measure, Spanish Reading CRT, English Reading CRT, Mathematics CRT, CTBS English Reading or CTBS English Mathematics. Pretests were administered in December, 1978.
Grade	Mean grade level of students taking dependent measure. The two grades represented in the sample are grades two and three.
Years Bil	Mean number of years of bilingual teaching experience of the teachers of students used in the analysis sample. The mean number of years for the entire group was two.

Program Variables

Imple. Qual.	An index representing how closely teachers conformed to SEDL guidelines for program operation as well as teacher perception of the quality of program components and their effectiveness. The index was constructed from the SEDL Implementation survey items 14a-g, 15, 16, 17, 18a-c, 20a, 20c, and 21. The index represents the standardized sum of responses to each item; the larger the sum, the more positive the quality of implementation.
Total time BSLC	Mean number of minutes of BSLC instruction in reading and mathematics.
Eng:Sp	Ratio of the total number of minutes of BSLC instruction in English to the total number of minutes of BSLC instruction in Spanish. The higher the ratio, the more instructional time was spent in English.

TABLE 6

Descriptive Statistics for SEDL Program Variables and Student Achievement Measures

		Criterion-Referenced Tests			Norm-Referenced Tests	
		Span. Reading	Eng. Reading	Mathematics	Eng. Reading	Eng. Mathematics
Posttest	$\bar{x}$	19.09	20.83	27.29	104.48	30.93
	SD	4.35	6.96	5.25	34.39	17.63
Control Variables						
Pretest	$\bar{x}$	16.73	17.05	22.52	89.40	28.98
	SD	4.60	7.28	7.35	36.27	11.58
Grade	$\bar{x}$	2.34	2.08	2.10	2.33	2.33
	SD	.59	.85	.88	.47	.47
Years Bil.	$\bar{x}$	2.56	2.23	2.45	2.07	2.07
	SD	.82	.84	1.50	.60	.60
Program Variables						
Imple. Qual.	$\bar{x}$	.86	3.56	6.24	.06	.06
	SD	13.77	9.94	10.69	3.79	3.79
Total time BSLC	$\bar{x}$	127.11	90.00	106.67	59.53	59.53
	SD	61.88	51.66	67.69	8.78	8.78
Eng:Sp	$\bar{x}$	.82	.61	.57	.57	.57
	SD	.42	.58	.47	.64	.64
N		131	267	228	97	97

Bilingual teaching experience was greatest for teachers administering the Spanish Reading CRT, followed by teachers giving the math CRT. The least experienced bilingual teachers had students who took the CTBS battery. Implementation quality was highest for the CRT mathematics teachers followed by the CRT English group, two of the more experienced teacher groups. Mean implementation quality was lowest for the CTBS teachers.

The amount of time spent in BSLC instruction was greatest for the Spanish CRT group, followed by the mathematics CRT group, the two most experienced bilingual teaching groups, and least for the teachers administering the CTBS English test. The range in mean number of minutes of BSLC instruction for the analysis sample is quite large; about twice as much BSLC instructional time was spent by teachers giving the Spanish reading test as by those administering CTBS tests.

The ratio of English to Spanish instruction, a measure of bilinguality of the program is quite interesting. The teachers who have the most bilingual teaching experience (2.56 years) and who spent the most time in BSLC instruction (127.11 minutes) also had the highest proportion of instructional time in English. These were also the teachers who administered the Spanish CRT, a finding opposite to what would be expected given the SEDL program goals and the teachers' backgrounds. The mathematics teachers, on the other hand, had the lowest proportion of English to Spanish instruction, indicating that they spoke more Spanish than their counterparts giving the Reading CRTs. These findings are important considerations in interpreting the relationship of bilingual teaching experience and time spent in BSLC instruction to student achievement discussed in the following section on Regression Results.

The intercorrelations among implementation variables are displayed in Table 7. Several findings are of interest here. The correlation of grade

TABLE 7

Intercorrelations Between SEDL Program Implementation Variables and Student Achievement Measures

		Criterion-Referenced Tests			Norm-Referenced Tests	
		Span. Reading	Eng. Reading	Mathematics	Eng. Reading	Eng. Mathematics
Control Variables						
Pretest	pre	--	--	--	--	--
	post	.63	.77	.56	.82	.54
Grade	pre	.43	.15	.26	.54	.64
	post	.26	.04	.15	.40	.44
Years Bil.	pre	-.03	-.35	-.05	-.38	-.16
	post	-.24	-.30	-.10	-.19	.05
Program Variables						
Imple. Qual.	pre	-.20	.12	-.24	.20	.01
	post	-.13	.23	.05	.23	.13
Total time BSLC	pre	.25	-.45	-.20	.39	.37
	post	.14	-.51	-.18	.27	.14
Eng:Sp	pre	.11	-.06	-.02	.13	.25
	post	.07	-.09	-.20	.21	.16
N		131	267	228	97	97

level to test performance decreases from pre- to posttest for all tests, indicating that posttest performance gains are not solely the result of maturation and student instructional history. The negative correlations between years bilingual teaching experience and the pretest scores suggest that the more experienced teachers were assigned students with the lowest English reading skills. These negative correlations of bilingual teaching experience with posttest results have several implications which will be discussed in the following section.

Correlations between quality of implementation and pretest scores indicate that teachers who were the "best implementers" were assigned the lowest achieving students in Spanish and the CRT mathematics group, while those "high implementers" administering the CTBS English reading test had slightly higher achieving students. For all groups of teachers, except those giving the math CRT, the correlation of attitude toward program and quality of implementation becomes positive from the pretest to the posttest period, suggesting that student gains are associated with positive teacher attitude.

Time spent in BSLC instruction is negatively correlated with English reading posttest performance. This student group had slightly lower pretest scores than the total sample (refer to Table 3 and 6) and were somewhat younger. Their mean ratio of English to Spanish instruction was second only to the group given the Spanish CRT. This group had lower English reading skills, but the English emphasis in BSLC instruction did not appear to address the language deficiencies of the group.

The negative correlation of the Eng:Sp index with mathematics CRT posttest performance reveals that for these students, performance increases as the amount of Spanish instruction increases and the proportion of English to Spanish declines. It is the opposite case for the relationship of language

of BSLC instruction to English CTBS reading; increasing the amount of English is associated with a slight increase in test score.

The intercorrelation results suggest that years of bilingual teaching experience, time spent in BSLC, and the ratio of English to Spanish instruction are important implementation variables that interact with student subgroup characteristics to produce differential outcomes in student achievement.

Regression Results for SEDL

Table 8 presents the results of the regression analysis relating SEDL students' achievement to measures of program implementation. The predictor variables used in the analyses are divided into two sets. The first set, control variables, includes variables whose influence should be controlled before evaluating the effects of the second set, the program variables. In the second set are the measures we constructed from teachers' survey responses to reflect the amount and quality of implementation of the SEDL-BSLC program. Note that we include the bilingual teaching experience variable, YrsBi, in the control variable set. The rationale for this is that we wish to examine the effects of the implementaion variables (i.e., the "program" set) without the possibility of confusing program effects with general effects of teaching experience. In other words, if only the experienced teachers implemented the program well, without controlling for YrsBi the analysis would mix the effects of implementation with those of teaching experience; the latter types of effects are likely to be present no matter what program the teachers happen to be working with.

Turning to the regression coefficients for the program variables, the first variable in the set, our implementation quality index, relates positively

TABLE 8

Regression Results for SEDL-BSLC Program[1]

	Criterion-Referenced Tests			Norm-Referenced Tests	
	Span. Reading	Eng. Reading	Mathematics	Eng. Reading	Eng. Mathematics
Control Variables					
Pretest	.695*[2]	.644*	.594*	.878*	.420*
Grade	-.115	.013	.180*	.040	.471*
Years Bil.	-.281*	-.044	-.144*	.476*	.804*
Program Variables					
Imple. Qual.	.010	.106*	.252*	.418*	.893*
Total time BSLC	-.084	-.212*	.156*	.043	-.012
Eng:Sp	-.016	.050	-.239*	-.005	-.177*
R^2	.460	.637	.408	.750	.565
N	131	267	228	97	97

[1]Coefficients are standardized partial regression coefficients.

[2]Starred coefficients are significant at the .01 level.

to achievement gain on all measures, with the effect being statistically significant in four out of five cases. The weak relationship for Spanish reading achievement may be due to the fact that teachers very rarely listed "strengthen or maintain Spanish language ability" as a primary educational goal (6%), in comparison to listing transition (31%) or bilingual (40%) instructional goals. The results for the implementation index, then, paint a consistently positive picture of program effects on student achievement.

The next program variable in the set, Total BSLC instructional time, suggests a more mixed picture of program effects. For three of the five achievement tests this variable makes no contribution to the prediction of performance. In the two instances where Total BSLC time is significantly related to achievement gain, the relationship is negative for the first (CRT English Reading) and positive for the second (CRT Mathematics). The only plausible explanation of the negative relationship between BSLC time and English reading, is that the teachers who spend larger amounts of time in BSLC instruction are working with students whose entry level English skills are severely limited, and therefore may be the most difficult students to work effectively in English with. Referring to the correlations in the previous table, we can see that this interpretation is partially supported; the correlation between Total BSLC time and the English CRT pretest is strongly negative, indicating that teachers who spend more time in BSLC have students whose initial English performance is extremely low. The positive relationship between Total BSLC time and CRT Mathematics performance probably reflects incremental gains associated with increased amounts of time working with the materials.

The final program variable included in the analyses, Ratio of English to Spanish instruction, should not be interpreted as a direct indicator of

program implementation. Rather, this variable tells us about the effects of different patterns of English/Spanish language BSLC instruction. Referring to the row of regression coefficients, it appears that Eng:Sp relates negatively, on the whole, with achievement gain. This is interpreted to mean that greater gains in BSLC instruction come about as a result of deemphasizing use of English in favor of Spanish as the language of instruction. Average values (see previous table of means and standard deviations) of Eng:Sp for the analysis samples range form .57 to .82 indicating that students are receiving more instruction in Spanish than in English.

Returning to the portion of the table showing the regression coefficients for the control variables, we should make some comment on the findings for the Years of Bilingual Teaching Experience variable. The pattern of results for this variable seems to be negative relationships to CRT test performance, and positive relationships to NRT test performance. On the surface this would suggest that more experienced teachers, while perhaps providing high quality instruction to their students, are less exclusively focused on the content of the SEDL materials, or are in some sense less rigorous implementers of the program. Instead, they may provide a broader range of instruction, perhaps at the cost of coverage of specific SEDL program objectives. Correlations among the independent variables used in the analyses lend plausibility to this interpretation. For each of the analysis samples except CRT Mathematics, YrsBi correlates negatively either with the implementation quality index (CTBS Reading and Mathematics, and CRT English Reading) or with Total BSLC time (Spanish and English CRT Reading). There appears to be little by way of corraboration for this interpretation of the negative finding in the CRT Mathematics analysis.

Another line of reasoning may help to clarify the negative finding for the bilingual teaching experience variable; we consider two additional pieces of information about YrsBi. Two final points concerning YrsBi should be mentioned. First, although each of the subsamples on which the outcome gain analyses were based contained teachers with comparable levels of bilingual teaching experience (averaging from 2.4-2.5 years), these teachers were somewhat less experienced than the total group of teachers who participated in testing (average = 3.2) and, although less strikingly, than the total sample of teachers returning surveys (average = 2.8). This relative lack of experience may make them a less than ideal sample on which to base inferences about experience in bilingual programs. Second, we know from another item in our implementation survey that only a fraction of those reporting experience in bilingual teaching also reported proficiency in the Spanish langauge*. This raises a question about what qualitatively, constitutes bilingual teaching, and suggests that perhaps teachers define this activity more broadly than was intended by the developers of the SEDL program.

*We have not actually crosstabulated the responses from these two items, but the frequency of reporting Spanish proficiency is sufficiently low to support the assertion.

SWRL Results

SWRL IMPLEMENTATION SURVEY

During the spring of each year of BSLC program operations, SWRL administered its Learning Center Accomplishment Summary (LCAS) to all teachers known to be in charge of Learning Centers. Even though the LCAS were designed primarily to provide SWRL with summary information essential for describing the BSLC Project to the Community Services Administration and other interested audiences, the survey also relates directly to the interests of the CSE evaluation. In keeping with the interagency arrangements for sharing data sources relevant to the BSLC program, SWRL made available to CSE copies of the data files containing LCAS survey responses. These data were analyzed by CSE in order to determine the extent and quality of implementation of the SWRL materials.

The purpose of this section is to present the descriptive analyzes of the LCAS survey data that parallel the SEDL implementation results obtained from the CSE-developed SEDL survey. We consider in this section only data obtained during the spring, 1979, administration of the LCAS, as the LCAS responses from the first year of program operation were presented in SWRL's seventh and eighth quarterly reports (June 30, 1978 and September 30, 1978).

Two samples were used in reporting the results of the SWRL evaluation, the total sample of BSLC Centers (N = 1479) from which an LCAS was returned, and the student testing sample (N = ranging from 582-716) including those second and third grade students participating in the program for whom Teacher Section sheets (cover sheets for CSE administered tests) were available. Although the difference in the number of centers represented in the total sample

(from on the LCAS) and in the analysis sample appears quite large, actual differences between them are often imperceptible. When teacher responses do differ by sample, these differences will be pointed out in the appropriate discussion section.

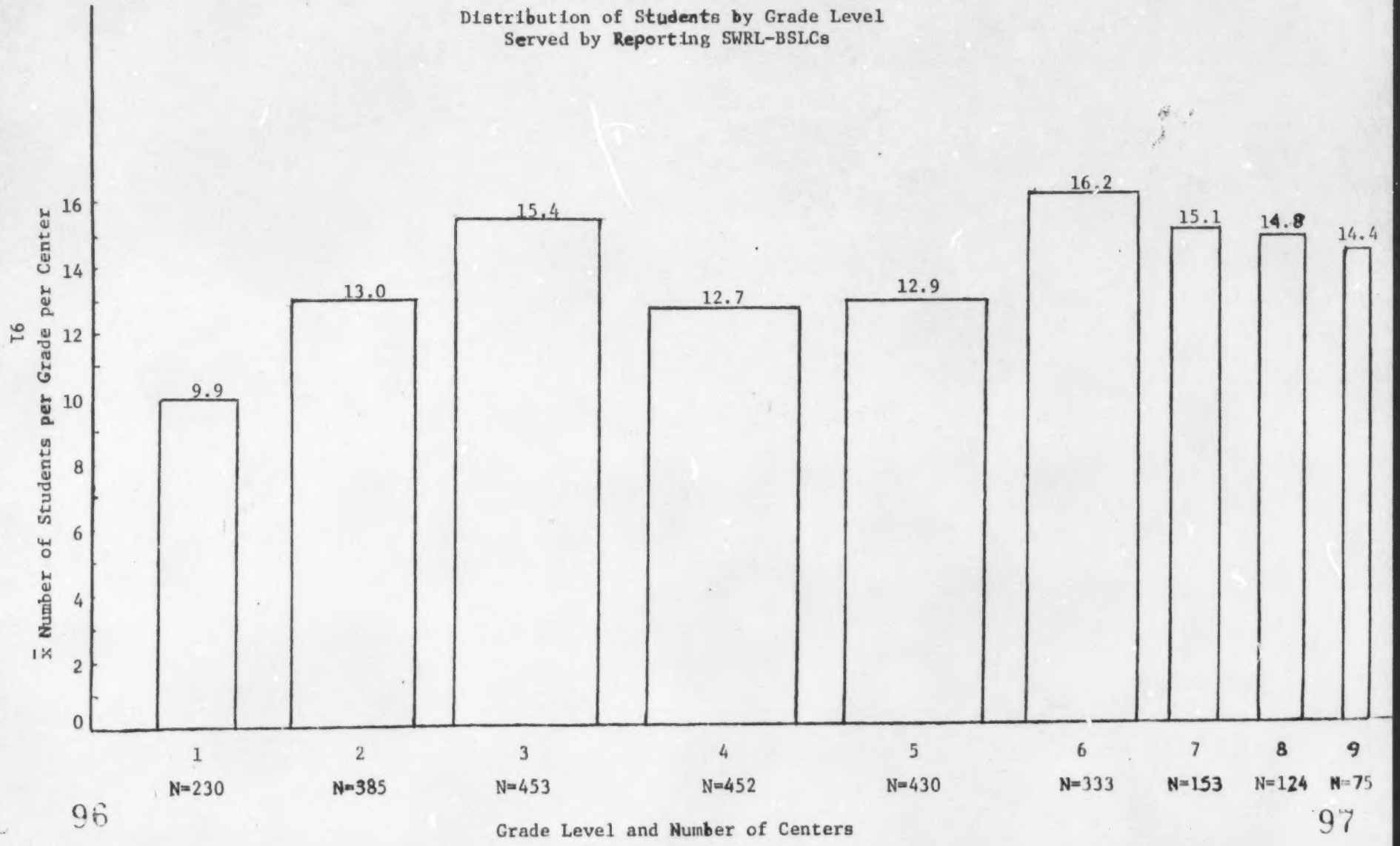
Description of Program Participants

SWRL-BSLCs served a wide range of grade levels, grades one through nine, in eleven states (refer to Appendix for location of SWRL sites returning spring posttest information and for summary information from the spring, 1979 LCAS). The mean number of students in each grade level participating in the BSLC program ranged from nine in the first grade to 16.2 in grade six; however, the mean number of pupils in each grade level was very evenly distributed. Figure 7 displays the mean number students in each grade level per center reported on the LCAS. The analysis sample was almost evenly divided between the second and third grades. Forty-one percent of the analysis sample was in the second grade and 49% in grade three (N = 668). From Figure 7 it can be seen that the sample of students selected for analysis of program outcome measures is representative in so far as number of students participating in BSLC instruction (385 to 453) as well as mean BSLC Center size (13 to 15 students).

Students served by SWRL materials did not appear to represent any one ethnic group to the degree that the SEDL materials, due to their bilingual target population, did. While particular BSLCs may have had disproportionate numbers of students classified as "ethnic minorities", the mean number of pupils at each center classified as American Indian, Black, Oriental, or Spanish surnamed was quite low in comparison to the average number of pupils served by a center. The mean number of American Indian was 3.2; Black, 3.0; Oriental, .25; and Spanish surnamed, 5.7, out of an overall mean across centers

FIGURE 7

Distribution of Students by Grade Level
Served by Reporting SWRL-BSLCs



of 23.7 pupils. Although Spanish surnamed pupils constituted by far the largest ethnic category, there is no indication from the data that BSLC participants included a large number of NES--LES students. The mean number of LES students reported was 1.8; a figure which may or may not represent Spanish-bilinguals.

Participating teachers overwhelmingly, were those who had been participating in the BSLC program for at least one year. Seventy-four percent of the teachers reported using BSLC materials last year.

Description of Program Setting

SWRL materials were used with almost equal frequency across a wide range of grades (1-9) and in regular classroom settings. Eighty-three percent of the respondents (N = 1486)* reported using the materials in a self-contained classroom environment. As a like percentage (84.8%, N = 406) reported adapting the materials to a regular classroom setting during the previous year's implementation, it is fairly safe to assert that the materials are routinely used in this fashion. Although there were no descriptive data to indicate how the remaining 17% of the teachers housed the BSLCs, 15% of the previous year's respondents reported that the materials were used in separate resource centers.

Assignment of Students to BSLC Program

The SWRL reading and mathematics programs are accompanied by a Placement Aid that enables teachers to determine which kits are appropriate for their students. A majority of the teachers took advantage of the aid to assign pupils to specific kits. Of 1142 teachers reporting their placement practices in reading, 52% used the placement aid most of the time and an additional 12%

*The N's in parentheses are the number of teachers responding to the item.

at least half of the time. Only 26% of the sample did not use the aid. The placement aide was used by a comparable number of teachers to assign students to mathematics materials, 53% (N = 973) reported using the aid almost always, and 10% in about half the cases. About one-quarter of the mathematics teachers did not use the aid at all, a figure comparable to the number not using the reading aid. Other procedures for assigning students to BSLC instruction teacher judgment (49% for reading, 43% for mathematics), other school testing results (26% for reading and 21% for mathematics), and teacher made tests. The LCAS does not indicate whether these latter methods were used in conjunction with or in lieu of the placement aid.

While the data for the total sample do not allow us to determine with much accuracy the percentage of students assigned to kits within grades, a review of the mean number of students assigned to each kit at each grade level (refer to BSLC-LCAS Summary, Appendix) indicates that, with the exception of the Decoding Skills Kits I-II, students were fairly evenly distributed across reading kits for the total sample. The distribution of students across mathematics kits suggests that more remediation was needed by students at the lower end of the math skill continuum (mean number of students = 11), BMS Kits 1-4, than the upper, where the mean number of students drops to 5 per kit.

The analysis sample, for which specific data related to kit placement were available, was not as evenly distributed across kits. More than half (52%) of the student were assigned to Basic Reading Skills Level I, and another 30% were placed in Decoding Skills I. The three Reading Development kits, I-III, accounted for less than 2% of the analysis sample. While the sample represented the lower end of the skill continuum in reading, this

was not the case with mathematics placement. More students were assigned to Basic Mathematics Skills, Level II (54%) than to Level I. Less than 3% of the students worked in BMS Levels IV and V and none were in Level VI, not an unusual finding for a predominately second and third grade group.

Use of Materials with Students

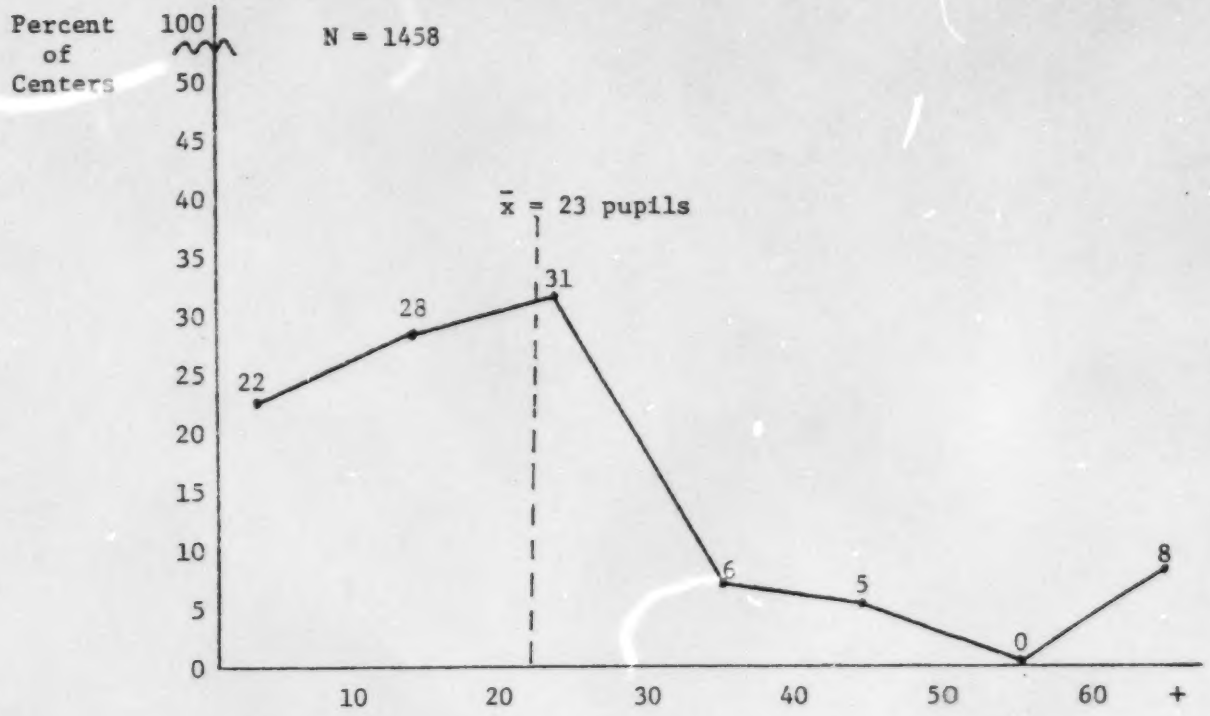
Although the LCAS does not provide information about the proportion of students assigned to the BSLC for remedial instruction, comparison of the mean number of students receiving any instruction in BSLCs with those for whom the materials were used only for remediation suggests that in many of the centers, the materials were not used exclusively for remedial instruction in reading and mathematics. Figure 8 displays the distribution of center sizes. The mean number of students receiving any BSLC instruction in the typical center was 23.7 (1479 centers reporting). The mean number of pupils for whom the BSLC instruction was remedial was 12.4 for reading and 10.0 for mathematics, with 1184 and 1029 centers reporting respectively. While no direct comparison of these figures can be made, the sheer size of the discrepancy between these means imply that many students included in BSLC programs were there for purposes other than remediation.

Coverage of SWRL Program Materials

Two indices of coverage of materials were included on the LCAS forms, one related to the speed of coverage and the second to the amount of materials covered during the school year. For the group of students assigned to BSLC reading instruction for remediation, within-kit progress was fairly slow. The mean number of students "not likely to complete all assigned kit sections" was 6.3; the mean number completing one kit was 7.9. These means outweigh the number of students either completing more than one kit (6.3) or ending BSLC instruction entirely (4.2).

FIGURE 8

Total Number of Pupils Receiving BSLC Instruction:
Distribution Across SWRL Centers



A similar progress pattern is evidence for mathematics remediation. The mean number of pupils not likely to complete all assigned sections of at least one BSLC Kit is 4.6, those completing assigned sections of only one kit average 6.6 pupils. These means are almost equal to the mean number of students completing work in more than one kit (6.4) or ending BSLC instruction (4.6). These progress categories are summarized in Figure 9. Progress in reading appears to be slower than in mathematics. A review of the number of weeks that the typical student takes to complete a full kit substantiates this observation. For the majority of reading kits, it takes 28 weeks to finish the material; for mathematics, however, completion time for the earlier kits is 26 weeks and for the last two, 24 weeks.

Completion rates for the mathematics materials appear to be higher for most of the kits, especially the earlier ones, than for the reading kits. Figure 10 displays the mean number of students completing each kind of BSLC kit. Given that completion time for the typical student working in BSLC mathematics is shorter by almost two full weeks than for the typical students working in reading, the higher completion rate in mathematics is to be expected.

The breadth of coverage for the entire sample is also indicated in Figure 10. The figure displays the mean number of students assigned to each kit in reading and mathematics. For the total sample, which represents nine grades and 167 to 625 centers, it can be seen that, with the exception of Decoding Skills I-II and Basic Mathematics Skills 5-6, students are fairly evenly distributed across kits. When the sample is analyzed by grade level, this finding does not hold, however.

Breadth of coverage for the analysis sample, grades 2-3, is displayed in Figure 11. This figure clearly shows that at the lower grade levels, more students are working in the early kits. It would be expected from these results

FIGURE 9

Profile of SWRL-BSLC Program: Mean Numbers of Students in Selected Progress Categories

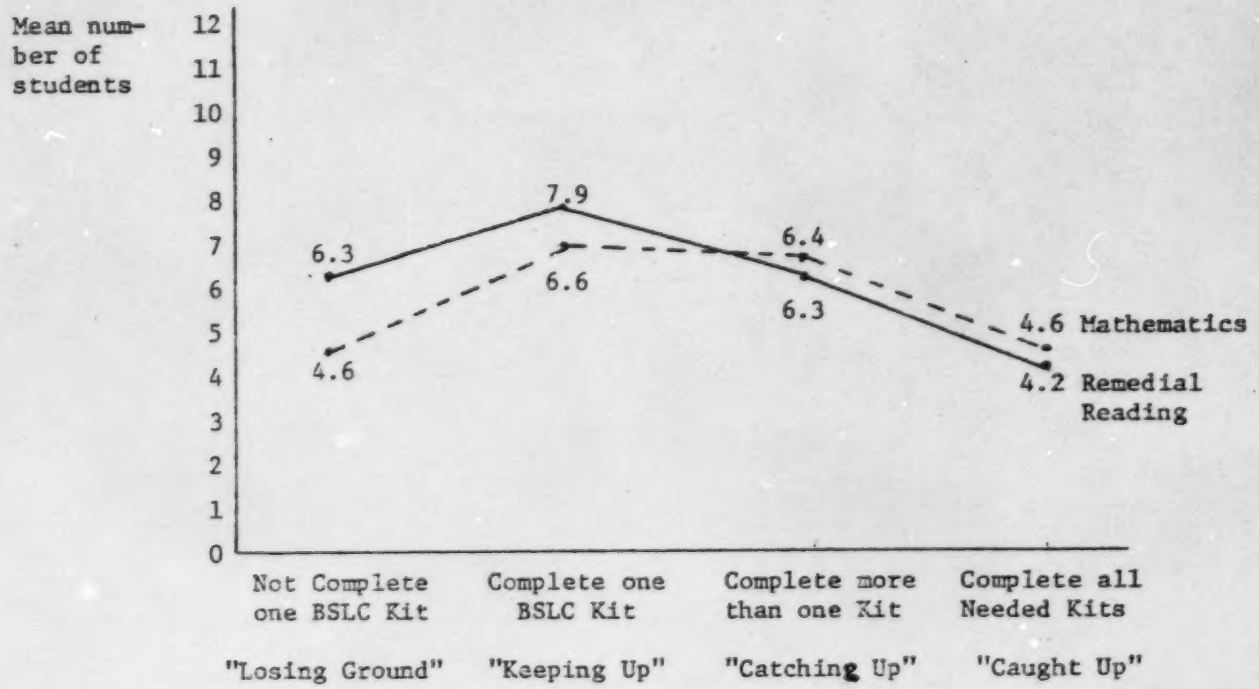
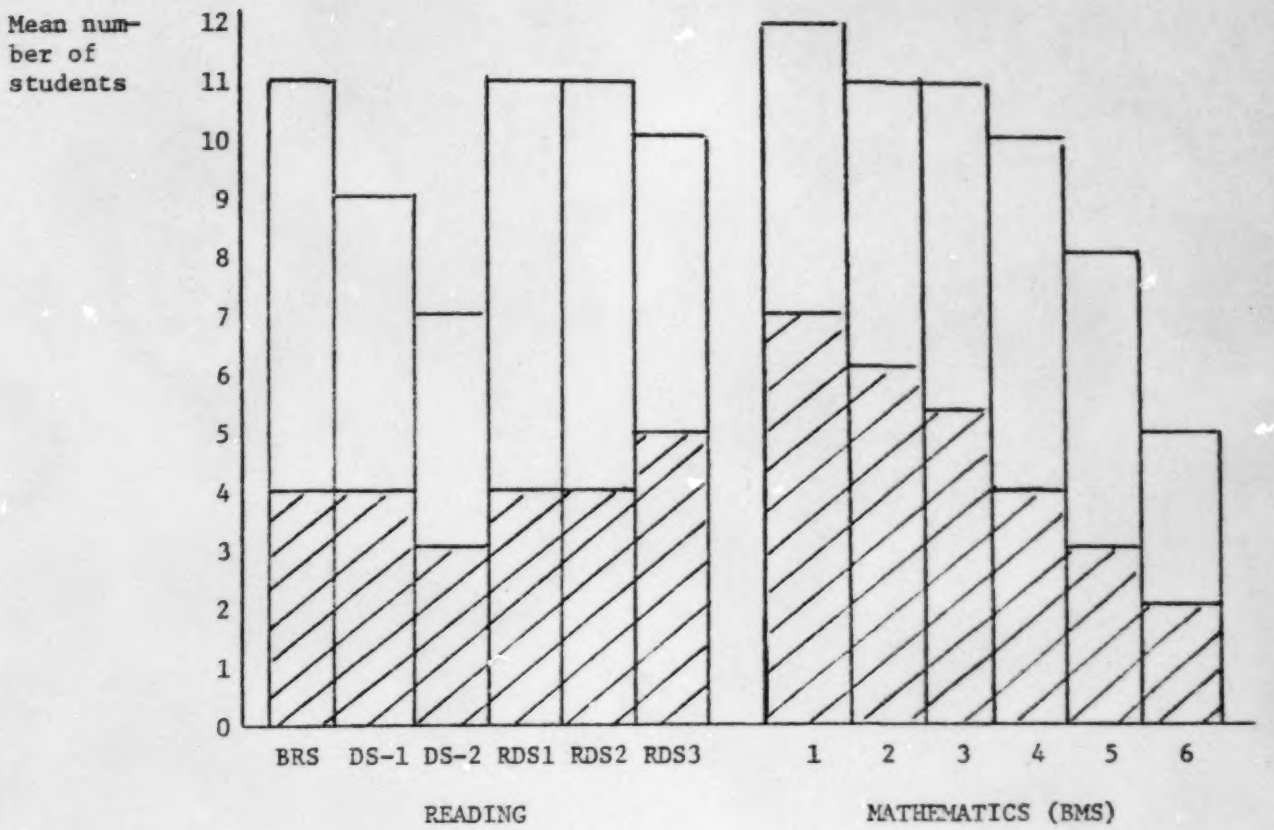


FIGURE 10

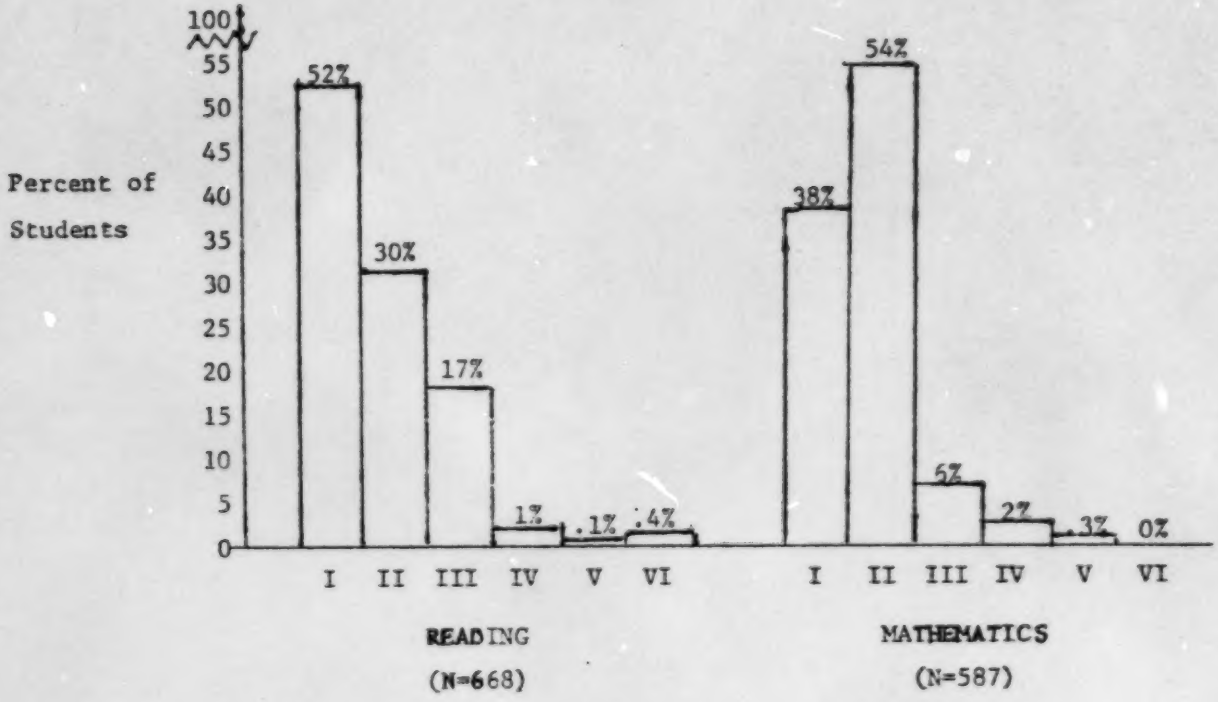
Mean Number of Students Completing each Kind of SWRL-BSLC Kit: Total Sample



Number of Centers represneted ranges from 167 to 620

FIGURE 11

Percentage of Students Working in SWRL/BSLC Kits
Spring: Breadth of Coverage
for Analysis Sample



that an analysis of breadth of coverage for students in the upper grades would display a larger proportion of students assigned to the higher level reading and mathematics kits; otherwise, the evenness in kit coverage evidenced by the total sample would not be possible (see Figure 10).

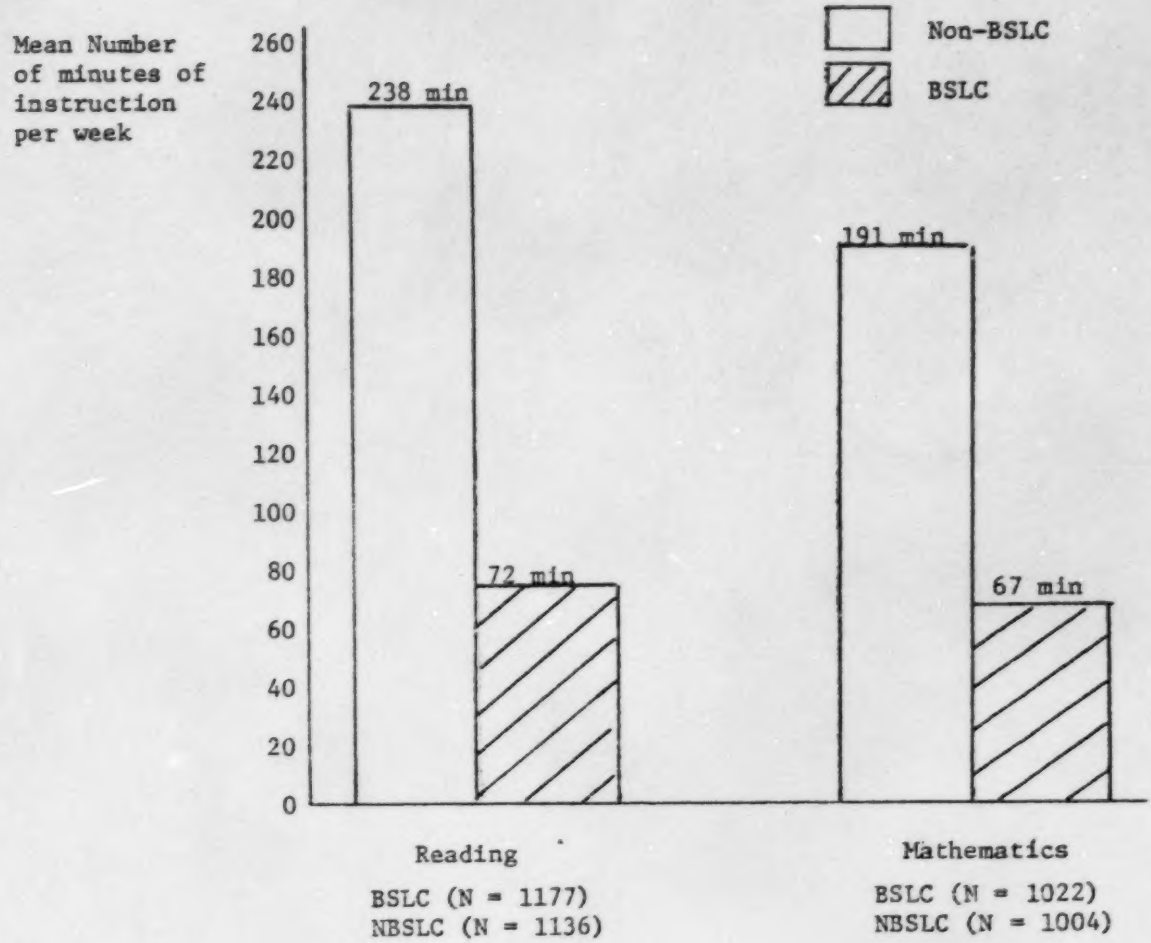
Time Spent in BSLC Instruction

BSLC v. NBSLC Time. The amount of time spent teaching reading and mathematics in regular instruction far exceeded the time spent in BSLC instruction for both reading and mathematics. Figure 12 displays the ratio of Non-BSLC to BSLC instruction for learning centers (n=1000+) and Figure 13 the ratio for students in the analysis sample. The ratio of Non-BSLC to BSLC reading instruction is smaller for the total sample than for the analysis sample, and very close to the ratio of Non-BSLC to BSLC mathematics instruction. For the analysis sample, about five times as much instructional time is spent in Non-BSLC reading than in BSLC reading, while the ratio of Non-BSLC mathematics to BSLC mathematics is 1.5:1. It is likely that both the discrepancies in the Non-BSLC to BSLC ratios for reading and mathematics, and in amounts of time spent in regular basic skills instruction versus BSLC instruction between the total sample and the analysis sample are explainable in terms of the differences in curricular emphases among grade levels. The total sample ratios summarize grades one through nine; the analysis sample, grades two and three.

The number of minutes per week typically spent in BSLC instruction for the total sample is displayed in Figure 14. At least one-half of the BSLCs reporting indicate that BSLC instruction is less than one hour a week, only 5-8% report spending more than two and one-half hours a week in BSLC tasks. The distribution of instructional time by percentage of centers reporting is very nearly the same for both reading and mathematics instruction. BSLC instructional time in reading is longer for the total sample than the analysis

FIGURE 12

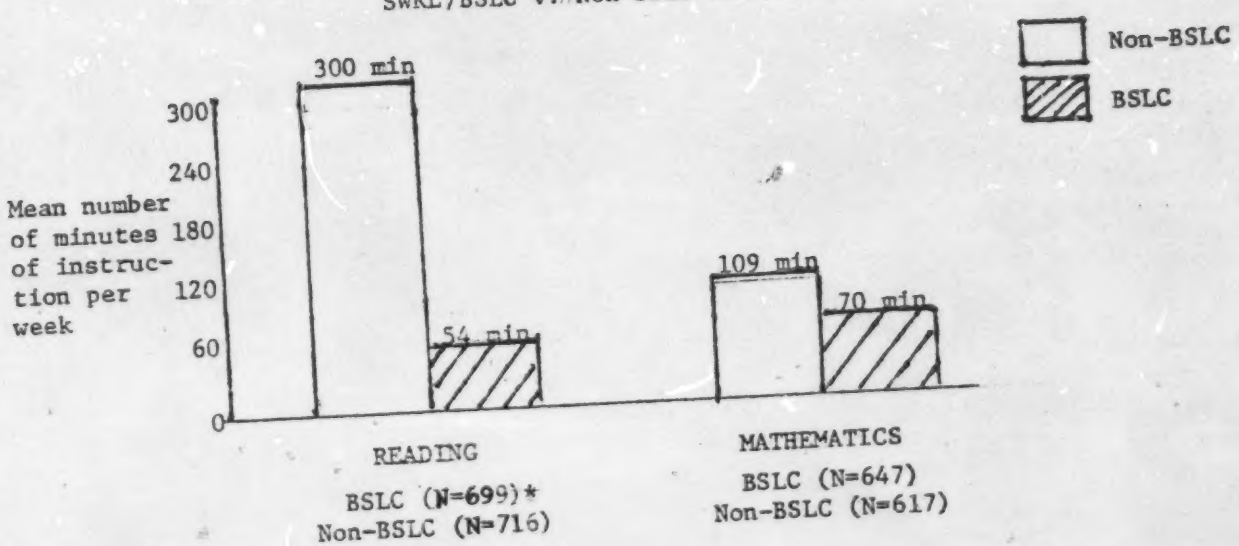
Mean Number Minutes per Week Spent in
SWRL-BSLC vs. Non-BSLC Instruction



Ratio of time of BSLC vs. NBSLC differs from total sample to analysis sample - due to grade level

FIGURE 13

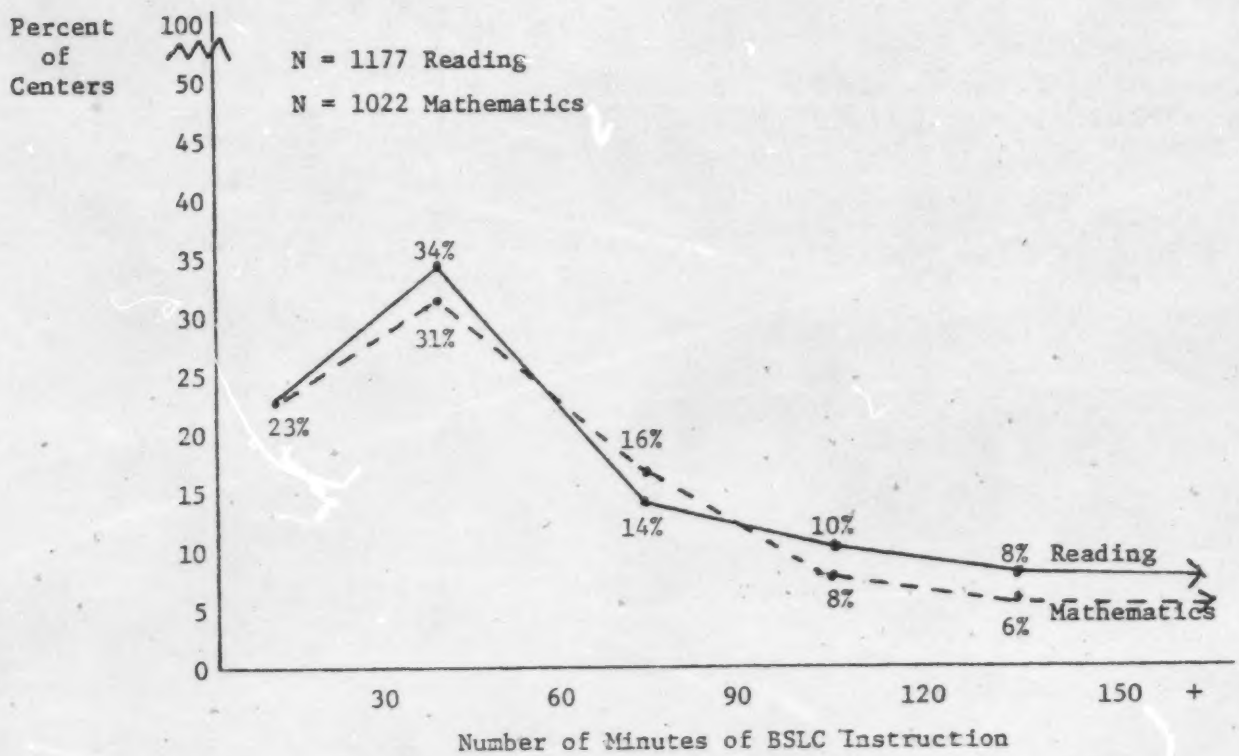
Mean Number of Minutes per Week in
SWRL/BSLC v. Non-BSLC Instruction



*n=number of students

FIGURE 14

Number of Minutes per Week Typically Spent by a Pupil in BSLC Instruction: Distribution Across SWRL Centers



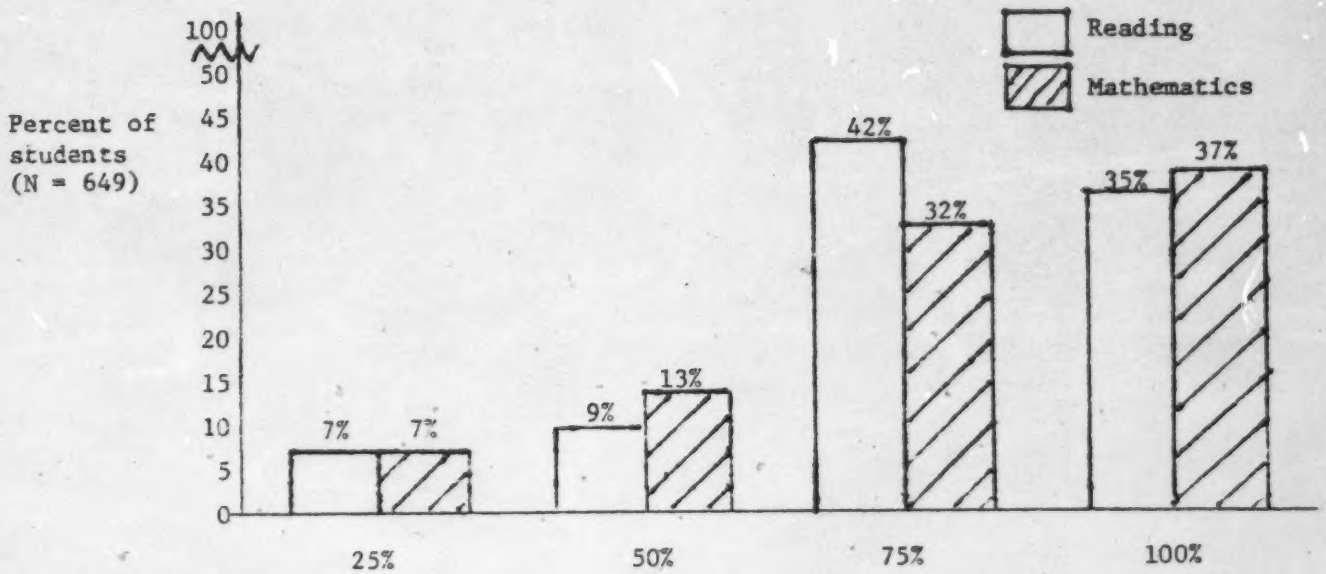
sample, 72 versus 54 minutes. The time spent in BSLC mathematics instruction for both groups is comparable, 67 versus 70 minutes. It must be remembered, however, that the mean times reported for the total sample are for centers, while those for the analysis sample are for individual students.

Time on Task. It is a fact of schooling that time devoted to instruction is not the same as time spent learning. In order to obtain a more accurate measure of the quality of BSLC instructional time, teachers were asked to indicate what percent of instructional time the students actually spent attending to learning tasks. The data for the analysis sample are displayed in Figure 15. For a majority of the students, at least 75% of the instructional time within BSLC was spent on task (n=649 for reading and 626 for mathematics). A surprisingly small number of teachers reported that individual students spent 25% or less of their BSLC time on task. The time on task comparisons indicate that SWRL materials were very effective in holding student interest.

Duration of BSLC instruction. Teachers supervising BSLCs were asked to indicate whether BSLC instruction for the year 1978-1979 ended earlier, at the same time, or later than it did for 1977-1978. Instructional duration of the BSLCs is a measure both of the time it took students to complete materials (discussed in the previous section) and of the time teachers were willing to donate from their own programs (84% of the BSLCs were located in regular classrooms). Compared to the previous year of implementation, the majority of the sample did not shorten the duration of BSLC instruction; 91% of the teachers supervising reading or mathematics centers responded that instruction ended at the same time or later than for the 1977-78 school year; 8% responded that BSLC instruction ended earlier in 1979 than in 1978.

FIGURE 15

Percent of Time Spent on SWRL/BSLC Tasks



* Mean time for reading was 53 minutes
Mean time for mathematics was 70 minutes

Teacher Attitude Toward Program

The Learning Center Accomplishment Summary included two sets of questions dealing with teachers' reactions to and attitudes toward the SWRL-BSLC program. The first set consisted of ratings of each materials kit on a one to five scale along the dimensions of interest to students, ease of use, comprehensiveness, and compatibility with teachers' educational goals. The second set of questions asked teachers to rate the operational and planning materials on each of the above dimensions except interest to students. Results for the kit ratings are summarized in Figure 16, and presented in detail in Table 9 . If we take the midpoint of the one to five range, i.e., a rating of three, as indicative of a neutral attitude toward the kits, it can be seen that all kits are rated positively on all dimensions. Ease of use ratings are the highest for both content areas, and these ratings increase for kits later in the sequences, probably reflecting the fact that older children can interact relatively independently with the materials in the later kits. The lowest ratings were assigned to the interest value the reading kits had for students. Otherwise, the ratings across kits were quite consistent.

The ratings for teacher operational and planning materials were also consistently positive (see LCAS item #19 in the Appendix). The highest rating was assigned to the compatibility of the materials (mean=3.6), with ease of use and comprehensiveness each being assigned ratings of 3.5.

Another measure of teacher attitude toward the SWRL-BSLC program is provided on the Teacher Section Cover Sheets for the analysis sample. Teachers are asked to indicate the effectiveness of either the reading program or the mathematics program for the individual student taking the SWRL CRT. Although, in general, teachers rated materials compatible with their educational goals, the reading program received an effectiveness rating of 1.8, where one

FIGURE 16

Profile of SWRL Kit Characteristics: A Summary of Mean Ratings Across Reading and Mathematics Kits

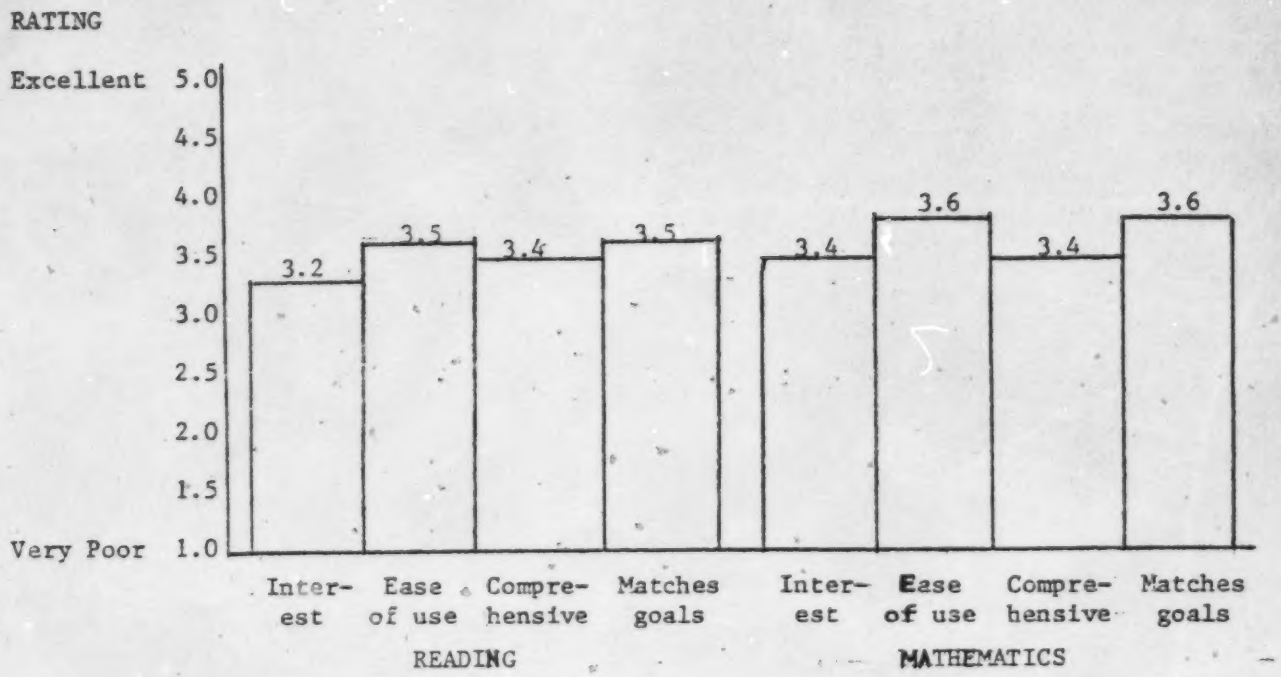


TABLE 9

Mean Ratings on Four Characteristics of the SWRL Reading and Mathematics Kits

Kit Characteristics	BRS	DS-1	DS-2	FDS1	RDS2	RDS3	BMS1	BMS2	BMS3	BMS4	BMS5	BMS6
Ability to capture pupils interest												
$\bar{x}$	3.2	3.2	3.2	3.3	3.3	3.3	3.4	3.4	3.4	3.4	3.3	3.3
S.D.	.87	.78	.83	.85	.85	.88	.87	.84	.87	.90	.97	.98
N	594	578	480	472	417	295	485	534	486	355	198	137
Ease of use												
$\bar{x}$	3.2	3.5	3.5	3.6	3.7	3.8	3.6	3.6	3.6	3.6	3.7	3.7
S.D.	1.1	.99	.97	.92	.90	.93	1.0	.92	.94	.96	1.0	1.1
N	593	574	476	470	416	292	486	534	485	353	197	135
Comprehensiveness												
$\bar{x}$	3.2	3.3	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
S.D.	.93	.90	.84	.86	.89	.92	.92	.88	.88	.88	.97	1.0
N	589	571	473	466	412	291	482	530	480	349	194	132
Compatability with educational goals												
$\bar{x}$	3.4	3.5	3.5	3.6	3.6	3.6	3.5	3.5	3.6	3.6	3.7	3.6
S.D.	.92	.91	.84	.89	.85	.90	.94	.91	.97	.94	1.0	1.1
N	586	573	475	469	416	294	483	534	485	354	199	134

Rating is as follows:

Very Poor				Excellent
1	2	3	4	5

represents a rating of "not effective" and five a rating of "very effective." By contrast, the mathematics program is rated 3.7 on the one to five effectiveness scale, a figure compatible with the ratings received by the individual mathematics kits on the dimension "compatibility with educational goals." The discrepancy between the effectiveness ratings for the reading materials and the quality ratings may be due to the selection bias in the teacher sample. Whereas the LCAS ratings were based on from 130 to 600 responses, the effectiveness ratings came from a sample of 80. Comparisons of the pretest scores in reading for the analysis sample with those of all SWRL students tested during the administrations from which the pretest scores were derived suggest that the former group was not an unrepresentative sample in terms of their achievement level. This comparison discourages us from interpreting the substantially lower effectiveness rating for the reading students in our sample as being based on differences between these students and those in other centers. Instead, the lower rating must be seen as an apparent discrepancy between corresponding pieces of information derived from different data sources.

SWRL STUDENT TEST DATA

Four student outcome measures were selected to assess student progress in the SWRL BSLC materials. Two of these tests were reading and mathematics tests jointly developed by SWRL and CSE to measure program specific objectives; the remaining two posttests were the California Achievement Test in Reading and Mathematics. (Refer to the section on instrument and development selection for a discussion.)

Psychometric Qualities of Outcome Measures

Reliabilities (coefficient alpha) were computed for each of the subscales of the four outcome measures. The Appendix displays the means, standard deviations

and alpha coefficients. Reliability for the SWRL Reading CRT ranged from .86 to .97; the SWRL Mathematics CRT had a reliability of .89. Reading subscales of the CAT had the following alpha coefficients: Phonic Analysis = .93, Structural Analysis = .78, Reading Vocabulary = .92, Reading Comprehension = .92, Spelling = .89, Language Mechanics = .89 and Language Expression = .91. The two CAT Mathematics subscales had reliabilities of .93 for Mathematics Computation and .93 for Mathematics Concepts. The descriptive statistics for the student achievement tests were calculated on Ns ranging from 917 to 1012 for the SWRL CRTs and 687 to 694 for the CAT. Item difficulty values (p. values) for each of the program posttests are reported in the Appendix.

As with the SEDL posttest, two levels of data analysis were performed. The statistical properties of the criterion measures were computed using the sample of students who had both pretest and posttest scores available. This sample, as noted above, ranged in size from 900 to 1000 for the criterion-referenced tests and from 680 to 700 for the norm-referenced measures. The sample upon which the analyses of the relationships between program implementation variables and student achievement were performed was considerably smaller. This smaller sample size (N = 78-80) was due to the matching procedure used to obtain data for the sample; thus students whose Teacher Surveys were incomplete or missing were deleted from the files.

Relationship of Implementation Variables to Student Achievement

Data in this study were collected for the ultimate purpose of determining which program implementation variables were positively related to gains in student achievement. As in the SEDL analyses, program implementation variables were constructed from the Teacher Survey data contained on the cover sheet of the SWRL posttest. The program implementation variables included two demographic variables related to the student sample, grade level and pretest score. The first is a measure of status or maturation and the second of background or prior instruction.

TABLE 10

Description of SWRL Implementation Variables

Control Variables

Pretest	Mean test score on SWRL Placement Aid, administered in the fall of 1977, or the SWRL End-of-Year test, administered in the spring of 1978.
Grade	Mean grade level of analysis sample. Range is grades two through six, with 49% of the sample in grade two and 41% of the sample in grade three.
Min.NBSLC	Mean number of minutes of non-BSLC instruction in reading and mathematics.

Program Variables

Min.BSLC	Mean number of minutes of BSLC instruction in reading and mathematics.
Time on Task	Index representing the percent of BSLC instructional time that students actually attended to learning. Range = 1-4, with 1 = 100% time on task, 2 = 75% time on task, and so on. As the variable is negatively keyed, higher scores mean that students are spending less of the time attending to the learning task.
Effectiveness	Mean rating of the effectiveness of either the SWRL reading or mathematics program for individual students taking SWRL CRT posttests. Effectiveness ratings range from one to five with one representing a "very ineffective" and five a "very effective" rating.

Another variable summarized teacher attitude toward the SWRL materials; this was the index of perceived material effectiveness. Teachers were asked to rate the effectiveness of the particular SWRL reading or mathematics materials for the individual student taking the posttest on a scale from 1 to 5. A rating of one is poor and five is very effective. The mean of student effectiveness ratings for materials associated with a particular posttest (SWRL Reading CRT, Mathematics CRT, CAT Reading or CAT Mathematics) provided the index of program effectiveness; underlying this index is the assumption that teachers who like the program will rate it as more effective than those who do not and the reverse.

Three implementation variables related to amount of BSLC and quality of BSLC instruction: Percent of time spent on task, Amount of time in Non-BSLC instruction and Amount of time in BSLC instruction. The percent of time on task variable is a summary index representing the mean percentage of students working "on task" in the materials 25%, 50%, 75% or 100% of the BSLC instructional period. The number of minutes spent in Non-BSLC and BSLC instruction is a straightforward mean of the minutes of instruction reported on the individual Teacher Sections of the posttests. Implementation indices were obtained from a sample ranging in size from 76-80 students.

Description of Implementation Variables: Descriptive Statistics and Intercorrelations

Multiple regression analysis (described in the first part of this chapter) were performed using the six implementation variables as independent, predictor variables for the posttests (dependent measures). The mean pretest scores reported for the SWRL Reading CRT and the CAT Reading, as well as those for the SWRL Mathematic and CAT Mathematics are the same because the same pretest, the SWRL Placement/End-of-Year Test, was used for taking

TABLE 11

Descriptive Statistics for SWRL Program Variables and Student Achievement Measures

		Criterion-Referenced Tests		Norm-Referenced Tests	
		Reading	Mathematics	Reading	Mathematics
Posttest	$\bar{x}$	34.157	21.119	109.451	55.611
	SD	7.788	5.065	15.972	6.244
Control Variables					
Pretest	$\bar{x}$	30.253	19.158	32.294	18.583
	SD	10.777	6.051	9.304	4.410
Grade	$\bar{x}$	2.916	2.737	2.961	2.889
	SD	0.280	0.443	0.196	0.319
Min. NBSLC	$\bar{x}$	249.570	184.868	225.882	191.667
	SD	94.721	65.084	94.094	80.835
Program Variables					
Min. BSLC	$\bar{x}$	61.446	47.303	62.353	53.472
	SD	33.745	42.461	38.915	50.154
Time on task	$\bar{x}$	1.651	1.842	1.549	1.444
	SD	1.839	1.120	1.119	1.054
Effectiveness	x	1.904	3.803	1.922	4.500
	SD	1.839	1.479	1.864	1.207
N		83	76	51	36

different posttests. For this reason pre- and posttest gain scores are not presented for the SWRL sample. Although the sample of students taking both a pre and posttest consisted predominately of third grade students (49%) and second graders (41%), the mean grade level for the mathematics tests was 2.7 as contrasted with 2.9 for the reading tests. The sample size for the analyses ranged from 78-80 students. The difference in mean grade level for the two groups taking the reading and math tests may reflect differences in the kinds of students assigned to each type of remedial instruction. Mean effectiveness of instruction is higher for mathematics than for readings; although students spent less time on task in mathematics than in reading. Both the total number of minutes of BSLC instruction and the total number of minutes non-BSLC instruction are greater for reading than mathematics by about 33%; the absolute value of time spent in non-BLSC reading and mathematics instruction, however exceeds BSLC reading and mathematics instruction by about 4:1.

Table 12 displays the intercorrelations among program implementation variables and the SWRL CRT and CAT mathematics and reading tests. The highest correlations with posttest are attained by the pretest measures, the SWRL Placement or End-of-Year Test. The next highest correlations for each of the outcome measures follow: CRT Reading correlates negatively with minutes of non-BSLC instruction and time on BSLC task. CRT Mathematics' second highest correlation is -.37 with time on task. Time on task also correlated negatively with CAT Reading. The CAT Mathematics posttest score correlates positively with both the index of effectiveness and breadth of coverage.

The amount of variation in posttest score explained by the predictor variables was greater for the CRT Reading test then the CAT (.47 vs. .40) but

TABLE 12

Intercorrelations Between SWRL Program Implementation Variables and Student Achievement Measures

		Criterion-Referenced Tests		Norm-Referenced Tests	
		Reading	Mathematics	Reading	Mathematics
Control Variables					
Pretest	pre	---	---	---	---
	post	.662	.485	.561	.628
Grade	pre	.185	.314	-.235	.108
	post	.236	.442	-.148	.150
Min. NBSLC	pre	-.253	-.016	-.362	.036
	post	-.277	.132	-.153	.174
Program Variables					
Min. BSLC	pre	.041	-.151	-.023	.131
	post	.019	.052	-.045	.188
Time on task	pre	-.178	-.103	-.018	-.070
	post	-.214	-.366	-.224	-.151
Effectiveness	pre	.089	-.247	.011	.067
	post	.063	.121	.150	.315
N		83	76	51	36

smaller for the CRT Mathematics test than the CAT (.45 vs. .52). Direct comparisons about relative contributions of the predictor variables to amount of posttest score must be made with caution, however, as the sample sizes upon which the regression squared correlations were obtained differed widely: The number of students for the CRT Reading was 83, for CAT Reading, 51. Ns for CRT Mathematics were 76 but for CAT Mathematics, 36.

Regression Results for SWRL

Table 13 presents the results of the regression analyses relating student outcome achievement scores in reading and mathematics to measures of program implementation. As with the corresponding analyses of SEDL data, we have again partitioned the independent variable set into control variables and program variables. The three measures in the program variable set are Minutes per Week of BSLC Instruction (Min.BSLC), "Time on Task", and Teachers' rating of the effectiveness of the SWRL program for each student (Effectiveness). It should be noted that the Time on Task variable is negatively keyed, so that higher scores on this variable mean that students are spending less of the assigned time attending to the learning task; correspondingly, coefficients indicating a positive relationship will be negative in value. It should also be noted in reviewing these results that the analyses cover a period of approximately one full calendar year, with the premeasure and the effectiveness rating obtained during the 1977-78 academic year, and the remainder of the variables derived from data gathered in the spring of 1979.

Considering the set of program variables as a whole, the results across the different outcome measures form a consistent positive pattern. Both Effectiveness and Time on Task are associated with improved student achievement

TABLE 13

Regression Results for SWRL-BSLC Program[1]

	Criterion-Referenced Tests		Norm-Referenced Tests	
	Reading	Mathematics	Reading	Mathematics
Control Variables				
Pretest	.588*[2]	.422*	.559*	.588*
Grade	.104	.214*	.049	-.117
Min. NBSLC	-.131	.099	-.026	.253
Program Variables				
Min. BSLC	-.010	.055	.111	.141
Time on task	-.085	-.322*	-.232	-.211
Effectiveness	.072	.110	.225	.157
R^2	.473	.446	.391	.502
N	83	76	51	36

[1]Coefficients are standardized partial regression coefficients.

[2]Starred coefficients are significant at the .01 level.

in both reading and mathematics measured by both CRTs and NRTs. Amount of BSLC instruction relates positively to achievement gain for three of the four analyses, being effectively unrelated to achievement on the CRT reading measure. Although the magnitudes of the coefficients are in all but one case not sufficiently large to achieve statistical significance, this should not be interpreted to mean the program effects are without practical significance. The squared multiple correlations (indicating the amount of the variance in the outcome measure explained by the predictor set) are all statistically significant.

Little in the results for the control variables requires comment. Minutes of non-BSLC instruction relates negatively to reading as measured by both types of tests, but positively to both CRT and NRT mathematics outcome scores. These findings suggest that the non-BSLC instruction SWRL students are receiving in reading is ineffective, but their non-BSLC mathematics instruction is contributing to their overall gains in mathematics.

CHAPTER V: CONCLUSIONS AND RECOMMENDATIONS

In the previous chapter on results of the CSE evaluation of the BSLC program, we presented a mass of information about the SWRL and SEDL programs and their impacts on the students they served. The purpose of this chapter is to draw out from that information those findings on which readers of this report should focus. These will include results that bear directly on the degree to which BSLC programs brought about positive change in the achievement of participating students; on the installation and use of BSLCs by teachers; and, on teachers' reactions to, and judgments concerning the worth of, the BSLC program. We present, for each BSLC program, these three sets of conclusions, followed by the recommendations we feel the conclusions warrant, along with any specific weaknesses in the evaluation that would cause us to qualify the statements made. The final section of the chapter outlines some general implications of our evaluation.

Conclusions and Recommendations Concerning the SEDL-BSLC Program

The CSE evaluation of the SEDL program has generated data and analyses that support a number of conclusions. Generally, we have found that the SEDL program serves a population of students whose status and levels of educational attainment conform to the pattern of characteristics identified by CSA. These students began their BSLC experience at a relative disadvantage vis-a-vis the students in the nation, and either maintained or improved their standing in terms of the various achievement areas covered in our battery of outcome indicators. The most important dimension of program implementation influencing student outcomes was the quality of teachers' perception and use of the BSLC program. Teachers gave the SEDL program high marks, rating program components positively with regard to effectiveness, comprehensiveness, and overall quality. By and

large, teachers in the sample received direct training from SEDL staff (68%) or from those receiving such training (20%). Despite this highly successful implementation of the training component, teachers lacked both the specialized teaching experience and language competencies needed for effective bilingual teaching; this, however, should not be seen as a fault of the SEDL program. These conclusions can be elaborated.

First, our analyses of the norm-referenced test scores of SEDL students produced mixed results. The several subtests of the CTBS can be divided into those measuring what can be regarded as rote learning or prerequisite skills, and those tapping more higher order skills such as comprehension and mathematics concepts. Interestingly, the areas where SEDL students advanced rather than maintained their relative standing were in the prerequisite skills which are both easier to teach and necessary for mastery of the higher order areas. This suggests that teachers are either correctly diagnosing the need for instruction in fundamentals and providing it effectively, or, given their relative lack of bilingual experience and biliteracy, are falling back on the easier areas to teach.

Second, the discrepancy between teachers' general reactions to the program and their levels of experience and appropriate teaching skills suggests that the implementation of the instructional component of the program was unlikely to have corresponded to the developer's classroom model. The favorable attitudes toward the program apparently are not grounded in the experience of "faithful" implementation of the SEDL program model. Instead, those teachers spending greater amounts of time in use of the materials did so through increasing emphasis on English language instruction, a pattern probably similar to their practices prior to adopting the SEDL program.

These findings in particular, and our understanding of the SEDL data generally, lead to some recommendations about future work with the SEDL program and other efforts to improve bilingual education. It is likely that the teachers in our survey sample are not unrepresentative of those involved in bilingual education throughout the United States. If this assertion is correct, it has implications for how bilingual programs should be disseminated and implemented. First, it is clear that the personnel-intensive approach employed by SEDL is not only wholely justified but may also need to be further intensified along specific lines. The program trainers may need to establish "qualifications" for potential adopters, and to formally build into the training program levels of use appropriate for teachers possessing particular levels of competence and experience. The SEDL training format already incorporates this notion informally through the conduct of introductory and advanced training sessions, but these sessions typically group users simply into those with and without prior SEDL experience. Also, it may be beneficial to articulate explicitly the relationships between patterns of teacher qualifications and patterns of student characteristics, so that the possibility of ineffective use is reduced. The general consequence of such modifications may be to limit the applicability of the SEDL program, but this limitation would more accurately reflect the readiness (instead of willingness) of the bilingual educational practitioner community to deal with the special educational problems of LEP students in new and effective ways.

Conclusions and Recommendations Concerning the SWRL-BSLC Program

CSE's evaluation of the SWRL-BSLC program has provided the information needed to develop a set of summary statements about the use and impact of the SWRL program. The preponderant use of SWRL placement procedures (the Placement Aid) by sample teachers provides assurance that BSLC students were in fact

initially in the remedial category. Rough indications from their outcome achievement levels are that they have advanced considerably in standing during their contact with the SWRL program. Students in the regression analysis sample were scoring above the 50th percentile at the end of the year of BSLC instruction, and variables measuring exposure to the SWRL program made positive contributions to the prediction of their outcome achievement status. Of particular importance in explaining SWRL growth was the proportion of the time in BSLC instruction they spent actively attending to the assigned learning task. The overall amount of time in BSLC instruction, and teachers' ratings of the effectiveness of the program for individual students made weaker but positive contributions to the prediction of achievement gain. Although the analysis samples were small, the availability of individual student level data on exposure to the program substantially increased the power of the analyses.

Teachers of students in the analysis sample rated the reading materials as somewhat less effective than the mathematics materials, but this difference was not corraborated in the responses from the total sample of SWRL teachers, who were uniformly positive in their ratings of the program. Teachers singled out the ease of use of the materials as a particularly positive feature of the program.

Regarding program usage, substantial proportions of students served by learning centers were classified by their teachers as using the materials in a remedial mode. Average numbers of students in each of SWRL's progress categories did not differ widely across categories. Most students were reported as "keeping up," meaning they were completing one SWRL kit per year, a reasonable rate of progress given teachers' estimates that roughly 26 weeks were needed to complete a kit. The students in our analysis sample were likewise approximately in the keeping up progress category, with reading students not quite

keeping up, and mathematics students doing slightly better than keeping up. Thus, the modal level of progress through the SWRL materials is, according to our outcome analyses, producing concrete achievement gains in SWRL-BSLC students.

The implications of the findings for the SWRL-BSLC program seem to be that their approach to providing remedial educational services is accepted and used by teachers, and that typical amounts of exposure to the program result in educational gains for participating students. Two qualifications need to be placed on the latter of these implications. First, as mentioned above, our outcome analysis samples were relatively small, making the results only "suggestive" rather than "conclusive" from a rigorous statistical standpoint. Second, we chose to focus in the SWRL evaluation on program implementation and impact at the low end of the operating grade range, and thus cannot speak with certainty about program effects outside the grade range we have studied. Our focus did, however, include the grade level at which the largest number of BSLC teachers reported serving students.

The BSLC Evaluation: Its Implications for the Field

A number of aspects of the CSE evaluation of the BSLC program stand out in retrospect as worthy of comment. In this section we discuss some of the events, experiences and perceptions that have accumulated over the three year life of the BSLC initiative in hopes that such discussion may help to guide future efforts at educational improvement and the future evaluation of these efforts. The intent here is not to express evaluative or judgmental views concerning the merits or impacts of the two BSLC projects. Three general topics are considered in the discussion:

- How the BSLC experience has affected our understanding of bilingual educational programs and bilingual evaluation.
- How the BSLC experience has affected our understanding of the issues in evaluating remedial educational programs.

- What are the strengths and potential pitfalls in large-scale, collaborative interagency efforts at educational improvement.

Obviously there is overlap among these topics; the grouping is intended to correspond to areas of concern to the distinct audiences likely to come in contact with this report.

Bilingual Educational Programs and Evaluations

Our experience with the SEDL component of the BSLC program lends itself to a number of observations relevant to bilingual programs. Several issues interact in the implementation and evaluation of a program such as that developed by the SEDL:

- Criteria for monitoring and placement of students
- Program implementation and the delivery of instruction in two languages
- District context and policies regarding bilingual programs in relation to program developers' objectives and funding agency requirements

We consider each of these in turn, both as they bear on the provision of educational services for bilingual children, and as they bear on the evaluations of such efforts.

Monitoring and Placement of Students. Programs purporting to be of general utility in the field of bilingual education must be sufficiently flexible to support instruction of students with widely different mixes of entering proficiency in Spanish and English. Self contained instructional packages that are developed by others than those directly responsible for teaching specific groups of students need to make explicit provision for the appropriate use of program components in the different conditions that will be found from class to class, district to district, etc. This will involve, first and foremost, the inclusion of procedures for identifying how students with particular skill profiles are to be assigned program components.

In assigning students to bilingual programs, it is often the case that students are assessed for language dominance as an entry level criterion. Dom-

inance tests indicate the child's relative strength in two languages, and would therefore seem useful in the determination of the appropriate language of instruction. One important, and often overlooked aspect of dominance tests is that the establishment of a student's dominant language does not necessarily give any indication of whether the student has sufficient skills in the dominant language to benefit from some given level of instruction in that language. Thus, at a minimum, assessment for dominance should be supplemented by some examination of the student's proficiency in the range of skill areas covered by the program, separately for the two languages.

Another issue in the assessment of multiple language skills in relation to effective prescription of bilingual instruction has to do with facilitative relationships between skills in two languages. It appears that the same level of dominance will have different implications for the language of instruction, depending on the level of proficiency of the student. Spanish dominant children who have limited proficiency in both languages may need to be instructed differently than Spanish dominant children possessing some degree of proficiency in Spanish. The former group may need to be taught basic concepts in their native language, whereas the latter group may already have acquired these concepts in their mother tongue and the transference to English will be facilitated by this condition.

For evaluations of bilingual programs, the issue shifts to ascertaining the appropriate language (or languages) in which to monitor student progress. Evaluations employing pre-post designs face the same set of problems as the teachers implementing the program when the evaluator attempts to obtain relevant and valid assessments of students' pre-program performance levels. The evaluation may need to employ a wide variety of methods for testing students, depending on the diversity of students being served by the program, and on the complexity (e.g., skill sequencing, individualization emphasis) of the program to be evaluated. An appropriately broad student assessment

battery will create additional problems for the data analyst on the evaluation team. The more measures that are employed, the smaller will be the sample sizes for any given instrument, and the weaker the inferences that can be based on a single instrument will be. Also, different forms of student assessment will have different psychometric properties that will need to be kept in mind when judging the degree of confidence that can be placed on findings.

Finally, large-scale evaluations of bilingual programs will usually need to depend on the active cooperation of the participants in the program, in particular, of teachers. In the context of student assessment, this means infringement on their instructional time. While one would expect teachers to be the strongest advocates of fair assessment of their students, the practical reality is that their contribution to the assessment process is influenced by factors that will not always lead to the best possible assessment of the accomplishments of their students. Teachers will wish to be seen as exemplars of their district's policies regarding bilingual students; if a district's orientation is toward transition, teachers will be less willing to argue that their students need to be tested in their native language.

Program Implementation and Instruction in Two Languages. The thrust of this issue for the developers and implementers of bilinugal educational programs relates to ensuring that users of the program have requisite competencies to implement it in an appropriate manner with the students for whom they intend it. Assuming the developer and user are not the same person (as is the case for the SEDL program), great care is needed in communicating (from developer to user) the skills and understandings required for effective use. The personnel intensive nature of the SEDL program is testament to this assertion. When the program under consideration is truly bilingual, effective use crucially depends on teachers ability to provide instruction in both languages. Without such ability, it can be stated with certainty that the program will

not be implemented as the developers intended.

The distribution of language skills among the teachers now responsible for the education of bilingual children in the United States must be regarded as an important unknown factor that will influence the effects of large-scale bilingual educational improvement efforts. For programs like SEDL's program dissemination occurs in waves: SEDL staff train those with the interests and resources to attend pre- and inservice sessions, who in turn train others. It is simply not known how many of those who make a decision to adopt such programs would be certified by the developers as competent to do so. The situation is a genuine paradox for those wishing to improve the lot of the LEP student. On one hand, program developers have good reason to disseminate what they know to be a beneficial practice. On the other hand, if the practice is adopted by teachers lacking the necessary skills, its impact on their students is sure to be diminished. Thus what is done in the service of bilingual education may not in fact alter the status quo.

Evaluators of bilingual program implementation face a complex challenge. Definition of "faithful" implementation of the program strictly in terms of the developers ideal will lead to the likely conclusion that no real change in practice has occurred in many adopting sites. Gauging implementation by the enthusiasm of program users in participating sites will undoubtedly bias the conclusion in the opposite direction; the sense of participation in a "movement" generates positive affect that may not be undergirded by actual improvements in practice.

The appropriate evaluation strategy for documenting bilingual program implementation will need to include both assessments of the fidelity with which users of the program carry out the developers intentions, as well as users attitudes and reactions to contact with the program. The question as to "when a program is really the program" is yet to be satisfactorily resolved; what is clear is that the answer is highly unlikely to be found through

survey research methods (e.g., questionnaires) alone. In addition, on-site observations of program use are required if the evaluators seek to fully document program implementation.

A second issue in evaluating program implementation concerns distinguishing sources of interference in the participating sites. For example, if implementing teachers cannot speak Spanish, this is a factor beyond the program developer's control. A related distinction is between the impact of the program in a particular set of sites (some of which may lack skills needed for effective implementation) and its general level of effectiveness under conditions deemed appropriate by the developers. One must of course go on to examine whether the conditions specified by the developer are realistic, but this is a separate evaluative task.

District Context, Developer Objectives and Funding Agency Requirements. Some of the issues in this section have been mentioned above; we highlight them here, and add to the list. Bilingual education has many of the characteristics of a "movement". Advocates use both political and educational rhetoric. This movement must be superimposed on existing educational institutions if it is to have the desired impact. The process of institutionalizing bilingual education occurs sporadically, being advanced by federal and state requirements, as well as by grassroots initiatives. Thus, the settings into which bilingual educational programs are placed will not all be uniformly receptive to such initiatives. One major "casualty" occured during the SEDL-BSLC program that must be understood in these terms.[1] To refer to losses of this nature as merely "attrition" misses an important dynamic.

The discussion above leads to the conclusion that, in the field of bilingual education, program developers work will be greatly facilitated

[1] We are referring here to the loss of a large district in the SEDL program brought about by changes in the educational administration of the district.

by shared understanding with the local sites who are their clients. The support for work such as the SEDL-BSLC program, however, derives from the sponsorship of federal agencies. This relationship entails a global view of the impact of the developers activities, so that the accountability of the federal agency is ensured. The evaluation, then, has an ultimate purpose of producing a simple and coherent summary statement about program impact; reconciling this purpose with the objectives and commitments of the developers and the clients they serve is a delicate undertaking. Since data to be gathered must inform a global question, its relationship to local needs, and the return from it for the time local participants invest in the evaluation may not be obvious.

Issues in Evaluating Remedial Educational Programs

One fact about remedial programs, and about most large scale interventions, is that they constitute a relatively weak "treatment" in the sense of occupying only a small portion of the total experience in the life of the targeted groups. The "problem" to be remediated, unlike the remedial program, cannot be measured in minutes per week. Instead, the problem of disadvantage has complex and persistent roots, largely outside of the schools. Even inside the classroom, remediation is, by definition, a supplementary activity, and not a radical alteration of the entire school experience. Pinpointing the influence on students of such an intervention requires a methodological sensitivity that is seldom attained in practice. In addition, looking in the wrong places for the effects (e.g., choosing an inappropriate indicator of impact) will almost certainly lead to null findings. From this perspective, much of the discussion and contention surrounding outcome assessment of remedial programs focuses on issues of immediate importance.

When we add to the problem of measuring the impact of remedial programs,

the program dimension of individualization of instruction, with the implicit absence of lockstep group progress through program content, designing a sensitive and comprehensive outcome assessment becomes an imposing task. In the case of a program like that developed by SWRL, where individual students guided by teachers' monitoring, can interact with the remedial materials at their own pace, there may be as many patterns of program impact as there are remedial students' entry level skill profiles. The applicability of a single outcome test to measure "appropriate" content is likely to be limited. On the other hand, tailoring the progress assessment to individual students is impractical in the context of a large scale third-party evaluation. (This point is not unrelated to the relatively small samples of SWRL students included in our regression analyses.) In addition, once the decision is made to "count" individual students' accomplishments as a method of outcome assessment the chain of assumptions that leads back to inferences about resolving problems of disadvantage is considerably lengthened.

In the BSLC evaluation, a concerted effort has made to strike a balance between a broad-based comparative assessment approach (NRTs) and gathering focused information about the achievement of specific program objectives (CRTs). In retrospect, the experience suggests the strategy may be a useful one. For both types of outcome assessment, however, careful analysis of the curriculum content is needed. The curriculum match for the program relevant instrument was achieved by incorporating items assembled by the developer to accompany the program materials. In choosing a normative instrument we matched candidate tests, item by item, with the content of the curriculum. The chosen test stacked up better in reading than in mathematics, which appears to be reflected in the relatively strong contribution of non-BSLC instructional time

to the prediction of outcomes on the mathematics portion of this test (see Table 13 in the Results Chapter).

Another "data driven" lesson, coming from the SWRL component of the evaluation, pertains to the extreme importance of measuring exposure to treatment variables at the individual student level. Particularly for programs placing emphasis on individualization, class level data will miss much of importance. Thus, even in small samples, when the phenomenon under investigation -- here, use of the program — is measured at the level at which it is occuring, the analyses possess sufficient power to uncover meaningful program-outcome relationships.

Strengths and Potential Pitfalls in Large-Scale Interagency Efforts

The BSLC initiative evolved from federal concern that remedial education be provided to disadvantaged students. The process leading to the developent of a federal program such as the BSLC project reflects a broad base of recommendations planning and analyzing program options. Often, as was the case in the BSLC program, the recommendations do not converge on a single strategy, but instead suggest multiple efforts that are either aimed at distinct groups within the target population or represent alternative approaches to accomplishing the agency's objectives. The evaluations of such multi-project programs are typically one of two types: (1) project specific evaluations conducted by project staff or under subcontract to them; and (2) independent, program-general evaluations, having the sponsoring agency as prime client. The former evaluation strategy may be seen for example in the Occupational Safety and Health Administrations programs to increase compliance with safety requirements, and the latter in the evaluation of the Follow Through Planned

Variation Program. Owing largely to the emphasis on interagency collaboration, the evaluation commissioned by CSA for the BSLC program attempted to synthesize these two approaches. Thus, the key features of the BSLC evaluation were the intent to tailor approaches to the specific components of the two BSLC projects (including instrumentation and field work), and at the same time retain a common core of evaluation procedures to support general inferences about the impacts of both projects. A number of generalizations can be culled from the experience of implementing such an evaluation:

- Understandings of program objectives and developer intents are gained in ongoing interaction that cannot be obtained by document reviews and analyses of program content.
- Collaboration among evaluators and program developers on methodological issues fosters a climate of confidence that design decisions are well reasoned and reflect the interests of each participating agency.
- The kinds of interagency agreements that form the basis for a smooth and responsive evaluation are of a different character than contractual provisions, and require skills and time commitments that are not usually considered dimensions of technical competence or line items in evaluation budgets.
- It is crucial to continuously monitor and document the understandings held by the different parties to the evaluation. Collegial rhetoric lends itself to interpretations of optimism and understanding, but conscientous effort is needed to transform such language into concrete guides to action.
- It is possible to formulate data gathering strategies that generate information useful for multiple purposes, i.e., for the evaluation and for program operation. Both the development of such strategies and the plans for the use of the data they yield require careful attention from the collaborating agencies.
- Frequent interaction and exchange among participating agencies creates a flexibility in program evaluation that is usually lacking. The BSLC evaluation benefitted immeasurably from this flexibility.